# GENDER BIASES IN EYEWITNESS MEMORY: RECALL AND RECOGNTION

A Thesis
by
MELISSA A. BAKER

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
MASTER OF ARTS

August 2015
Department of Psychology

GENDER BIASES IN EYEWITNESS MEMORY: RECALL AND RECOGNITION

A Thesis
by
MELISSA A. BAKER
August 2015

APPROVED BY:

Paul A. Fox, Ph.D.
Chairperson, Thesis Committee

Twila A. Wingrove, Ph.D.
Member, Thesis Committee

Robert W. Hill, Ph.D.
Member, Thesis Committee

James C. Denniston, Ph.D.
Chairperson, Department of Psychology

Max C. Poole, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

**Abstract**

GENDER BIASES IN EYEWITNESS MEMORY: RECALL AND RECOGNTION

Melissa A. Baker
B.A., Appalachian State University
M.A., Appalachian State University

Chairperson: Paul A. Fox, Ph.D.

This 2 (participant gender: male, female) × 2 (criminal gender: male, female) × 3 (timing of the witness description: immediately, 1-4 days later, or not at all) factorial study examined gender differences in eyewitness memory for both recall and recognition. Participants viewed a brief breaking and entering video and 1-4 days later attempted to identify the criminal in a 6-person simultaneous lineup. Previous research has supported issues relative to cross-racial identification to the extent that juries in some states now require jury briefing concerning that effect. Mixed results have been found relative to cross-gender identification. Contrary to the female superiority main effect, my results indicated that females were not significantly better at describing (recalling) or identifying (recognizing) criminals in lineups compared to male eyewitnesses. However, female eyewitnesses did provide longer descriptions of criminals compared to male eyewitnesses. Inconsistent with the own-gender bias, interactions between participant gender and criminal gender showing that eyewitnesses were better at describing and identifying criminals who were of their own gender were not found. As expected, there was no relationship between eyewitness'

descriptions of the criminal and the accuracy of their identifications. When recall was examined as a function of time of recall after viewing the crime video, no differences were found between the immediate recall group's and the delayed recall group's descriptions of the criminal. However, when identifications were examined as a function of the time after which participants were asked to recall a criminal's description after viewing a crime, improved identification accuracy was observed for the delayed recall group compared to the immediate recall group.

*Keywords:* own-gender bias, female-superiority effect, eyewitness memory, recall, recognition, description accuracy, descriptor quantity, identification accuracy, time of recall

## Acknowledgments

I would never have been able to finish this thesis without the guidance of my committee members, help from various faculty members and research assistants, and financial support from the Psychology department and university.

I would like to express my deepest gratitude to my advisor, Dr. Paul Fox, for his excellent mentorship, support provided throughout the development of my thesis, and corny (but much appreciated) jokes. I would like to thank thesis committee member, Dr. Twila Wingrove (aka T$), for patiently correcting my writing, her willingness to work with me, and allowing me to call her names other than the one on her driver's license. I would also like to thank committee member Dr. Robert Hill for his cheery disposition and time spent reviewing and revising this thesis.

I would like to acknowledge Program Director, Dr. Rose Mary Webb, for her formatting guidance according to American Psychological Association and Cratis D. Williams Graduate School standards. A special thanks goes to Dr. Verne Bacharach for his help throughout my undergraduate and graduate studies and for assisting me in my understanding of advanced statistics; many statistics of which were essential for this thesis. I would like to thank research assistants, Bethany Poff and Rachel Krauza, for assisting me with data collection.

Lastly, I would like to thank the Psychology Department's WiSE Research Award program and Appalachian State University's Office of Student Research Grant and the Cratis D. Williams Graduate Research Grant programs for funding this research project.

## Dedication

I dedicate this thesis to my parents, Nancy and Allen Baker, who have always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve. This thesis is also dedicated to my best friend and sister, Mary, who has been a constant source of support and encouragement during the challenges of graduate school and life. I also dedicate this thesis to my beloved dog, Suey, for companionably choosing to nap at my feet during the writing of this thesis. Lastly, I dedicate this thesis to those innocent persons who have been wrongfully tried and convicted based on mistaken eyewitness testimony.

## Table of Contents

## List of Tables

## List of Figures

## Foreword

This thesis is written in accordance with the style of the *Publication Manual of the American Psychological Association (6th Edition)* as required by the Department of Psychology at Appalachian State University.

# Gender Biases in Eyewitness Memory: Recall and Recognition

Melissa A. Baker

Appalachian State University

Abstract

This 2 (participant gender: male, female) × 2 (criminal gender: male, female) × 3 (timing of the witness description: immediately, 1-4 days later, or not at all) factorial study examined gender differences in eyewitness memory for both recall and recognition. Participants viewed a brief breaking and entering video and 1-4 days later attempted to identify the criminal in a 6-person simultaneous lineup. Previous research has supported issues relative to cross-racial identification to the extent that juries in some states now require jury briefing concerning that effect. Mixed results have been found relative to cross-gender identification. Contrary to the female superiority main effect, my results indicated that females were not significantly better at describing (recalling) or identifying (recognizing) criminals in lineups compared to male eyewitnesses. However, female eyewitnesses did provide longer descriptions of criminals compared to male eyewitnesses. Inconsistent with the own-gender bias, interactions between participant gender and criminal gender showing that eyewitnesses were better at describing and identifying criminals who were of their own gender were not found. As expected, there was no relationship between eyewitness' descriptions of the criminal and the accuracy of their identifications. When recall was examined as a function of time of recall after viewing the crime video, no differences were found between the immediate recall group's and the delayed recall group's descriptions of the criminal. However, when identifications were examined as a function of the time after which participants were asked to recall a criminal's description after viewing a crime, improved identification accuracy was observed for the delayed recall group compared to the immediate recall group.

## Introduction and Literature Review

### Gender Biases in Eyewitness Memory: Recall and Recognition

It is estimated that eyewitness evidence has played a role in over 75,000 court cases per year in North America (Goldstein, Chance, & Schneller, 1989; Wells & Seelau, 1995). More innocent citizens are wrongfully tried and convicted on the basis of eyewitness evidence than by any other factor within the legal system (Innocence Project 2010; Memon, Gabbert, & Hope, 2004; National Institute of Justice, 1999, 2003; Wells & Seelau, 1995; Wright & Davies, 1999). Studies suggest that eyewitness accounts are the primary evidence used by prosecutors and are the most sought-after form of evidence during the investigative process (Brigham & WolfsKeil, 1983). Jurors typically regard it as the most useful evidence in a trial (Boyce, Beaudry, & Lindsay, 2007; Brigham & WolfsKeil, 1983; Lindsay, 1994; Lindsay, Lim, Marando, & Cully, 1986; Lindsay, Wells, & O'Connor, 1989). Because convictions are frequently based on eyewitness testimony, courts (Innocence Project 2010; Pezdek, 2012; Wells, Greathouse, & Smalarz; 2012) have begun to re-examine the reliability of such testimony. Recently, the jurisprudential system seems to be becoming more aware of the fallibility of eyewitness testimony and of the plethora of research that has been conducted that can provide a foundation for more skeptical consideration of its reliability in court.

In *Neil v. Biggers* (1972), the United States Supreme Court established a two-pronged approach to determine whether eyewitness testimony should be allowed in a specific case. The court ruled that the trial judge should determine whether the questioning of the witness by the police was unnecessarily suggestive, and if it was, was it nevertheless reliable. For the finding of reliability, the Supreme Court indicated that the following five questions must be addressed: 1) did the witness have adequate opportunity to view the scene; 2) was the time between the incident and the identification inappropriately long; 3) how certain was the

witness in his or her identification; 4) to what extent did the witness attend to the scene of the incident; and 5) how accurate was the eyewitness' initial identification. The *Neil v. Biggers* ruling was upheld by the Supreme Court in *Manson v. Braithwaite* (1977).

More recently in *State v. Henderson* (2011), the New Jersey Supreme Court ruled that the *Neil v. Biggers* standards fail to provide an adequate measure of "reliability," do not deter police misconduct, and rely too much on jury competence to evaluate reliability. The New Jersey Supreme Court reviewed more than 200 scientific articles on eyewitness testimony and ruled that when there is evidence of suggestiveness, all relevant system variables and estimator variables should be examined. Among those variables that affect eyewitness identification accuracy, a system variable is one that is (or could be) under control of the criminal justice system, while an estimator variable is one that is not. System variables include instructions given to eyewitnesses prior to viewing a lineup and the functional size of a lineup. Estimator variables include lighting conditions at the time of witnessing and whether the eyewitness and criminal are of the same or of different races (for more on system and estimator variables, see Wells, 1978).

Similarly, in Oregon, *State v. Lawson* (2012) concurred with *State v. Henderson* and gave the state the burden to prove the eyewitness conditions meet the admissibility standards. *State v. Lawson* also allowed expert testimony in unclear cases of identification, and the judge can disallow questionable testimony relative to scientific evidence. The new legal framework requires Oregon courts to consider all of the factors that may affect an identification's reliability and instructs courts, where appropriate, to employ remedies, such as limiting the witness's testimony and permitting expert testimony to explain the scientific research on memory and identification. The Court was particularly concerned with the effects

of suggestion (and in particular police suggestion) on memory and likened identification evidence to other forms of physical trace evidence, finding that "it is incumbent on courts and law enforcement personnel to treat eyewitness memory just as carefully as they would other forms of trace evidence such as DNA, bloodstains, or fingerprints, the evidentiary value of which can be impaired or destroyed by contamination" (*State v. Lawson*, 2012).

My study was designed to examine estimator variables that might affect eyewitness testimony. Specifically, I wanted to investigate cross-gender effects in eyewitness testimony. In doing so, I assessed the accuracy of eyewitness descriptions and whether a cross-gender effect existed in eyewitnesses ability to describe a criminal. The time between the opportunity of participants to view a brief breaking and entering video and a request for a description of the criminal was varied in order to determine whether the time delay influenced the description accuracy. My study also assessed the accuracy of eyewitness identifications. Based on the evidence of difficulty with cross-racial identification, an examination of whether a cross-gender effect exists in eyewitnesses ability to identify criminals was conducted. In addition, I sought to examine the relationship between the accuracy of an eyewitness's description of a criminal and the eyewitness' subsequent identification decision. Identification accuracy as a function of the time after which eyewitness' were asked to provide a description of the criminal was also assessed.

**Eyewitness Descriptions as Recall Memory**

Asking an eyewitness to describe a criminal is a recall task. Unlike recognition, recall is the retrieval of information from memory without a cue (Kahana, Rizzuto & Schneider, 2005; Payne, Elie, Blackwell, & Neuschatz, 1996). Recall involves remembering a fact, an event, or an object that is not currently physically present (in the sense of retrieving a

representation, mental image or concept) and requires the direct retrieval of information from memory (e.g., remembering a person). For example, instead of looking at a lineup, an eyewitness must describe the person they saw to a sketch artist. The sketch artist may try to help eyewitnesses' recollection by asking questions, but ultimately eyewitnesses have to retrieve the information themselves.

Information obtained from eyewitnesses is typically assigned great importance in criminal investigations. In a survey of British police officers, Kebbell and Milne (1998) found that more than one-third of the respondents agreed that eyewitness statements "always" or "almost always" provide major leads for an investigation, while another 51% claimed that such is "usually" the case. Not surprisingly, the important role of eyewitness reports has spurred a great deal of research investigating witnesses' memories for criminal events (Ross, Read, & Toglia, 1994), including eyewitness identification of offenders (Sporer, Malpass, & Kohnken, 1996). One largely neglected topic of research, however, is the process of recalling and describing criminals (Sporer, 1996), and thus relatively little is known about the performance of eyewitnesses in this regard.

Although few in number, archival studies have provided fairly consistent findings concerning the most general properties of criminal descriptions. For instance, a single eyewitness typically describes 10 attributes of the criminal (Sporer, 1996). Moreover, the most frequently mentioned attributes seem fairly similar across studies (Kuehn, 1974; van Koppen & Lochun, 1997). These generally include gender, age, height, and build—that is, basic information conveying a relatively vague portrait of the criminal. In cases in which the criminal does not wear a disguise, however, facial features are commonly reported, with the most frequent details being facial shape and skin descriptors (Sporer, 1996).

A few studies have compared the information reported by eyewitnesses with the actual appearance of the criminal (Farrington & Lambert, 1997; Sporer, 1996; Tollestrup, Turtle, & Yuille, 1994; van Koppen & Lochun, 1997). Although these studies have indicated that criminal descriptions are usually vague and general, information concerning salient aspects of a criminal's physical appearance—such as gender, ethnicity, and facial features—tend to be highly accurate. For instance, in both the Farrington and Lambert (1997) and the van Koppen and Lochun (1997) studies, reports of these particular descriptors were completely or partly accurate for more than 80% of the eyewitnesses. Furthermore, eyewitnesses were fairly precise in their descriptions of age, build, height, hair color, and hairstyle.

However, as argued by several psychologists, archival assessments of eyewitness accuracy should be treated with caution (Davies, 1992; Farrington & Lambert, 1997; Sporer, 1996). In these studies, researchers relied on police records when determining the accuracy of criminal descriptions. Several sources of error are associated with such procedures. First, as Sporer (1996) pointed out, one rarely knows for sure who actually committed a crime. That is, one cannot fully dismiss the risk that the person depicted in the police file and the person described by the eyewitness are not the same. Second, Davies (1992) addressed the problem of incomplete data associated with archival studies. That is, accuracy scoring is possible only for criminal attributes that are known to the police (e.g., gender). Thus, important information about a particular criminal (e.g., race) may simply not be verifiable. Third, some attributes that are in fact recorded by the police are likely to have changed since the time of the crime (e.g., clothes, haircuts, facial hair). Finally, data found in police files may be flawed simply because of unreliable documentation procedures. Indeed, Farrington

and Lambert (1997) found, when examining their verification data, that some features (e.g., voice, hair color, height) were unreliably or incompletely documented for one-third or more of the criminals. A more recent archival study by Fahsing, Ask, and Granhag (2004) found similar results. Their study was conducted using 250 offender descriptions by eyewitnesses of armed bank robberies. The accuracy of their descriptions was compared to authentic surveillance videos of the crimes. They found that eyewitnesses provided accurate descriptions of the criminals but reported few identifying details.

Unlike previous archival studies, this study is not limited to secondhand sources of information for verification of criminal descriptions. Instead, I complemented previous research studies (e.g., Meissner, Sporer, & Susa, 2008; Pigott & Brigham, 1985; Pozzulo and Warren, 2003) by gauging descriptions against video footage of the witnessed events. For example, while not conducted using a forensic paradigm, Pigott and Brigham (1985) conducted one of the first studies to assess the accuracy of eyewitness descriptions. In their study, 120 participants viewed a target person for 15 seconds and later described the target person's physical characteristics. Pigott and Brigham found that participants generally provided accurate descriptions of the target. Like Pigott and Brigham, Pozzulo and Warren (2003) conducted another study assessing description accuracy. After their participants viewed a criminal in a brief video, they were asked to describe that criminal. Pozzulo and Warren (2003) also found that participants provided accurate descriptions of the criminal. While these studies were not conducted using forensically relevant eyewitness paradigms, such studies do provide insight about persons' ability to describe accurately another person's physical appearance. Studies that have been conducted in a forensic setting also help us understand eyewitness description accuracy. For example, Meissner et al. (2008) presented

the first meta-analysis that examined several eyewitness measures including description accuracy, across 33 research papers. The studies used similar eyewitness paradigms—the same design that was employed in my study. Participants ($N$ = 4,278) viewed brief videos of a crime and were asked to describe the criminal. Results suggested that eyewitnesses are accurate at describing criminals via a simulated crime paradigm.

Combined findings from both archival and research studies suggest that, overall, people are accurate at describing other people's physical characteristics. While the archival studies provide important real-life data, experimental support provides data obtained from controlled experiments. Based on this previous research, I expected to find similar results in my research—participants will be accurate at describing the criminal.

To enhance consistency with the literature and external validity, methodology for my study will be similar to the forensic eyewitness paradigms used in the research cited above. Participants viewed a video of a crime and were asked to describe the criminal later. The value of using a video to enhance the methodological validity of real-life studies was recently acknowledged and demonstrated by Woolnough and MacLeod (2001). Hence, my research provides a methodological improvement over previous studies in at least three ways. First, I am able to evaluate descriptions of details that are not normally reliably documented in police records (e.g., weapon, disguise, and clothes). Second, I can be sure that the descriptions were compared with the actual appearance of the criminals as per the time of observation. Third, the likelihood that eyewitness descriptions would be matched against the appearance of innocent suspects is eliminated.

**Does Gender Affect Description Accuracy?**

As described in the introduction, I investigated the impact of two estimator variables on eyewitness memory. The first estimator variable was gender. Past research suggests that women outperform men in various recall tasks (Herlitz, Nilsson, & Bäckman, 1997; Herlitz & Rehnman, 2008; Lindholm & Christianson, 1998; Zelinski, Gilewski, & Schaie, 1993). Research suggests that women also outperform men when verbal content is used in recall tasks (Lewin, Wolgers, & Herlitz, 2001; Loftus, Miller, & Burns, 1987). These gender differences are said to exist as a result of women's superior verbal abilities, which contribute to greater recall (Herlitz & Rehnman, 2008). Given this consistent pattern of female superiority in recall tasks, I expect the same pattern to exist in eyewitness descriptions.

As expected, research results for eyewitness testimony have shown a female advantage in the number of details and accuracy of recall (Areh, 2011; Lindholn & Christianson, 1998). Lindholn and Christianson (1998) investigated how gender of the criminal is related to the eyewitness's description accuracy of the criminal. In their study, 80 male and 84 female participants watched a film depicting either a male or female criminal committing a simulated murder. Results revealed a female advantage in overall description accuracy of both male and female criminals. In addition, Areh (2011) conducted a study investigating sex differences in the accuracy of an eyewitness' recall for a criminal. In their study, 280 participants (58% female, 42% male) watched a 2 minute video of a robbery. Later, participants were asked to describe the criminal's appearance. Results revealed that female participants outperformed male participants in the accuracy of the descriptions. Based on the abundance of research suggesting a female recall advantage, I expected to find a similar female superiority main effect in my study. I hypothesized that female participants

would outperform males in the accuracy and quantity of descriptors of the criminal (*Hypothesis 1a*).

While research provides support for a female superiority main effect in description accuracy, other studies suggest that there may be more to gender differences related to description accuracy. A study by Powers, Andriks, and Loftus (1979) found that men and women were better at describing people of their own gender, which would be consistent with an own-gender bias interaction. In their early but compelling study, Powers et al. (1979) conducted two experiments assessing gender differences in eyewitness accounts. In their study, 250 participants (50% female, 50% male) viewed a series of slides depicting a wallet snatching (Experiment 1) or a fight (Experiment 2). After viewing the crimes, participants were asked to recall the incident. They found that female participants were more accurate when describing the female criminal and male participants were more accurate when describing the male criminal.

The Powers et al. (1979) study appears to be the only one to support own-gender bias in an eyewitnesses description of criminals. More empirical support is needed to determine whether an own-gender bias exists with respect to eyewitness recall. Because of the paucity of empirical support in recall studies, an own-gender bias for eyewitness descriptions will not be hypothesized in this study, but will be explored (*Hypothesis 1b*).

**Does the Timing of the Description Affect Its Accuracy?**

There is little empirical research regarding how the time after which eyewitnesses are asked to recall a criminal's appearance after viewing a crime (i.e., immediately or delayed) might influence the accuracy of the description. I sought to fill that gap in the literature. One study provides indirect evidence about whether the timing of the description influences

description accuracy. For example, Wixted and Ebbesen (1997) had participants view a simulated crime and asked them to write (recall) details of the crime before picking the criminal from a lineup one to five days later. They found that the time between the crime and participants' recall of the crime (approximately 1-5 days) did not affect the accuracy of their descriptions.

While Wixted and Ebbesen's (1997) findings suggest little loss of memory despite time of recall, memory experts agree that accuracy of recall decreases with time (Kassin Tubb, Hosch, & Memon, 2001). In eyewitness memory research, these experts have described diminished memory as a rapid early decrease, followed by a leveling off. For example, Laney and Loftus (2009), in their review of the eyewitness literature, concluded that there is sufficient evidence that eyewitnesses' memory loss follows the forgetting curve of Ebbinghuas (Ebbesen & Rienick, 1998). This theoretical curve illustrates memory as being highly accurate soon after the initial experience followed by a decline with the passage of time. The forgetting curve, of course, is the curve that specifies the strength of the memory representation over a particular retention interval. Based on that rationale, I expected to find that memory will be best soon after viewing the crime video and will weaken after time has lapsed. Specifically, I expected to find that participants who are asked to recall immediately will have more accurate descriptions of the criminal compared to participants who were asked to recall after a delay or not at all (*Hypothesis 2*).

**Eyewitness Identifications as Recognition Memory**

Recognition is a response to a sensory cue (Kahana et al., 2005). When one sees something, one compares it to information stored in memory, and if one finds a match, one recognizes it. Recognition is a largely unconscious process, and the brain has a dedicated

face-recognition area (Payne et al., 1996). Because recognition is the association of an event previously experienced or encountered, it involves a process of comparison of information with memory (e.g., recognizing a face). A police simultaneous lineup is a classic exercise in recognition. In a simultaneous lineup, a witness looks at several people and compares each person in the lineup to the person they "remember" committed the crime.

Eyewitnesses are critical in solving crimes, and sometimes eyewitness testimony is the only evidence available for determining the identity of the culprit. Psychological researchers since the 1970s, however, have consistently articulated concerns about the accuracy of eyewitness identifications. Using various methodologies, such as filmed events and live staged crimes, eyewitness researchers have noted that mistaken identification rates can be surprisingly high and that eyewitnesses often express certainty when they mistakenly select someone from a lineup.

Much of what we know about identification accuracy comes from archival data. Several such studies have been conducted, both in the United States (Behrman & Davey, 2001) and in the United Kingdom (Horry, Memon, Wright, & Milne, 2012). The results of archival studies have been inconsistent and, at times, contradictory. Tollerstrup, Turtle, and Yuille (1994) published the first archival study of eyewitness identification, sampling all cases of robbery and fraud that took place within a given time period in one Canadian police force. They found that victims were not very accurate in identifying suspects. Other British archival studies (Pike, Brace, & Kynan, 2002; Wright and McDavid, 1996) found similar results. Findings from these studies suggest that eyewitness identification accuracy is, overall, poor. However, in contrast to Tollerstrup et al. (1994), Pike et al. (2002) and Wright and McDavid (1996), archival data from Scotland suggests that eyewitnesses have a high

identification accuracy rate (Memon et al., 2011). In addition to archival data, evidence for low eyewitness identification accuracy is also mirrored in the experimental research literature. For example, in their meta-analysis, Meissner at al. (2008) examined several eyewitness measures including identification accuracy. After viewing a crime, participants engaged in an identification task. Results suggest that eyewitnesses are not very accurate at identifying criminals from lineups.

Across the archival studies and research reports described, the support for poor identification accuracy outweighs the support for good identification accuracy. However, when focused on the influence of certain estimator variables, eyewitness identification accuracy might change. For instance, much of the research on eyewitness identifications has sought to examine eyewitness characteristics that affect eyewitness reliability.

Specific research has been done to examine whether members of certain groups are better eyewitnesses than those of other groups. For example, the age of the eyewitness has been consistently linked to eyewitness identification performance, with very young children and the elderly performing significantly inferiorly to younger adults when the target is a younger adult (i.e., own-age bias). In addition to age, the race of the eyewitness has been examined extensively. Studies have shown that when the eyewitness is of a different race than the suspect, accuracy rates are lower than in same-race identifications (i.e., own-race bias; Brigham & Barkowitz, 1978; Brigham, Bennett, Meissner, & Mitchell, 2007; Bothwell, Brigham, & Malpass, 1989; Chance, & Goldstein, 1996; Cross, Cross, & Daly, 1971; Malpass & Kravitz, 1987). One type of own-group bias that has received relatively little attention is an own-gender bias; that is, whether males may be better at recognizing other

males than females and whether females may be better at recognizing other females than males.

**Does Gender Affect Identification Accuracy?**

Just as there is a clear pattern of evidence of female superiority in description accuracy, there also appears to be female superiority with regard to recall accuracy relative to eyewitness identification (Areh, 2011; Lewin & Herlitz, 2002; Rehnman & Herlitz, 2007). For example, Lewin and Heriltz (2002) showed participants either a male or female target face. Later, participants were asked to identify the target from a lineup. Results show that female participants were more accurate in identifying both male and female faces from the lineup. Studies by Rehnman and Herlitz (2006, 2007) also examined gender differences in facial recognition. Rehnmen and Herlitz (2006) had 109 female participants and 88 male participants view faces for later recognition. Results showed that female participants outperformed male participants in recognition of both female and male faces. An additional study by Rehnman and Herlitz (2007) had 107 men and 112 women view faces for later recognition. As in their previous study, women were especially good at recognizing female faces, but also outperformed men on male faces. These studies demonstrate a female superiority main effect in identification accuracy. Based on this research suggesting a female recognition advantage, I expected to find a similar female superiority main effect in my study. I hypothesized that female participants would outperform male participants in identification accuracy of the criminals of both genders (*Hypothesis 3a*).

Beyond a female superiority effect, there might also be an interaction effect when it comes to gender-related eyewitness identification. Specifically, there is some evidence to suggest that women might be better at identifying female faces, while men might be better at

identifying male faces. This phenomenon has been labeled "own-gender bias" and parallels the "own-race bias" that has been consistently demonstrated in the psychological literature (Bothwell et al., 1989; Brigham & Barkowitz, 1978; Brigham et al., 2007; Chance, & Goldstein, 1996; Cross at al., 1971; Malpass & Kravitz, 1987).

Suggestion of an own-gender bias emerged in recognition studies outside of the eyewitness literature. For example, several studies conducted in the 1970s and 1980s suggested that men outperform women in recognizing male-oriented objects, but women outperform men in recognizing both female-oriented objects and neutral objects (Loftus et al., 1987; McGivern et al., 1998; Powers et al., 1979). Researchers have also investigated whether there is an own-gender bias in the accuracy of facial recognition. Shapiro and Penrod's (1986) meta-analysis of 128 recognition studies, some of which manipulated the gender of face and gender of participant, found support for an own-gender bias; men were better at recognizing male faces and women where better at recognizing female faces. However, the own-gender bias was smaller than own-race bias results, and the size of the own-gender bias varied across the studies.

One potential reason for this variability is that several of the studies included in the meta-analysis looked at recognition memory for faces outside of the eyewitness context (Chance, Goldstein, & McBride, 1975; Cross at al., 1971). Two studies that were done after Chance et al.'s (1975) meta-analysis were conducted by Shaw and Skolnick (1994, 1999). The researchers looked at the possibility of own-gender bias in recognition memory for faces within the context of eyewitnesses. While they found partial support for own-gender bias in recognition memory for faces, it failed to find full support for such a bias (Shaw & Skolnick,

1994, 1999). These results suggest that the earlier meta-analysis findings may not extend to eyewitness scenarios.

In addition to the question of whether own-gender bias exists, there is an open question about what the nature of the interaction would look like. If the own-gender bias mirrored the own-race bias, men would be better at identifying male faces and women would be better at identifying female faces. However, this resemblance to the own-race bias has not been observed for own-gender bias. For example, Shapiro and Penrod (1986) found that the bias does not appear to occur to the same extent for women and men. In particular, the own-gender effect has been more consistently demonstrated for women than for men in studies conducted outside of forensic contexts (Areh, 2011; Cross at al., 1971; Jalbert & Getting, 1992; Lewin & Herlitz, 2002; McKelvie, 1987; Rehnman & Herlitz, 2007). On the other hand, Wright and Sladden (2003) found that the strength of the effect was the same for both women and men. Still, other studies have found that men are better at recognizing women (Brown, Deffenbacher, & Sturgill, 1977; Witryol & Kaess, 1957; Yarmey, 1993), which is inconsistent with an own-gender bias.

A major aim of my study was to examine gender differences in eyewitness testimony. By combining all of the research on gender differences in eyewitness identifications, it is unclear whether an own-gender bias exists. I investigated whether an own-gender bias existed in this sample. Given the inconsistent support for the presence of an own-gender bias, especially in the forensic context, I did not make a specific hypothesis (*Hypothesis 3b*). However, Figure 1 illustrates what the data would look like if both an own-gender bias and the female superiority effect were found.

**Does Engaging in Recall Affect Identification Accuracy?**

The second variable that I investigated was the role of recall, or description by the eyewitness, on identification accuracy. When the accuracy measure is a recognition task—eyewitness identification—there are three questions related to how prior recall might affect that recognition. First, for those who engage in a verbal recall task (eyewitness description), is the accuracy of their recall related to the accuracy of their subsequent recognition (eyewitness identification)? Second, do people who are required to perform a recall task perform differently on a subsequent identification task than those who are not? Third, does the timing of the recall impact identification accuracy?

**Is description accuracy related to identification accuracy?** With regard to whether there is a relationship between description accuracy and identification accuracy, the U.S. Supreme Court has explicitly endorsed the belief that there is a meaningful and useful relationship between the two (*Neil v. Biggers*, 1972). A past review of the experimental literature, however, calls this belief into question (Wells & Murray, 1983). Specifically, research seems to indicate that people who are superior at describing details of faces from memory are not appreciably superior at recognizing faces (Goldstein, Johnson, & Chance, 1979; Howells, 1938; Wolfskiel & Brigham, 1985).

An early study by Goldstein at al. (1979) tested the Court's (*Neil v. Biggers*, 1972) implication that people who are “good describers” are also “good identifiers.” Specifically, their participants were shown faces that they had to describe from memory and were later shown a different set of faces that they had to recognize from memory. In other words, the faces participants described were not the ones they had to recognize, which is clearly

different from a criminal eyewitness situation. Nevertheless, their research is not totally irrelevant. They found no relationship between description and identification accuracy.

Wolfskiel and Brigham (1985) conducted one of the earliest direct tests of the "good describer-good identifier hypothesis." They exposed participants to one of two target persons under conditions in which participants made deep or shallow judgments of the target at the time of encoding. Later, participants gave verbal descriptions and then attempted to identify the target person among distractors. Their results indicated that participants who gave relatively accurate descriptions were not more likely than those who gave relatively poor descriptions to identify the actual target.

A large body of research (Flexser & Tulving, 1978; 1982) demonstrates that persons who can recall a word when provided a retrieval cue often fail to recognize the word on a recognition test. This general finding, based on several variations in paradigm (Tulving & Thomson, 1973), has been termed *retrieval independence*, suggesting that cues that are present on a recognition task are uncorrelated with those on a recall test. Counter to the Supreme Court's assertion, accuracy of prior descriptions (recall) is not likely to be related to the ability to identify someone from a lineup (recognition).

More pertinent to my study, research investigating the relationship between recall accuracy and description accuracy using a forensic eyewitness paradigm has also been conducted. A series of studies by Schooler and Engstler-Schooler (1990) demonstrated that the act of generating a verbal description might do considerable harm to an eyewitness's memory for a criminal when he or she is asked to make an identification from a lineup. In particular, Schooler and Engster-Schooler asked participants to view a 30 second crime video of a robbery, followed by a 20 minute distractor task. Participants were then randomly

assigned to one of two conditions. One group was asked to write a description of the criminal, while the second group was assigned an unrelated activity. Immediately following those activities, participants engaged in the identification task. Their results indicated that participants who described the robber (38% accuracy) were significantly less accurate on the identification task compared with participants in the no-description condition (64% accuracy). In five subsequent studies, Schooler and Engster-Schooler repeatedly demonstrated this effect, known now as *verbal overshadowing.*

While subsequent research has largely confirmed the findings of Schooler and Engster-Schooler's (1990) verbal overshadowing effect, and the empirical literature has been summarized in several reviews (Meissner & Brigham, 2001; Schooler, Ryan, & Reder, 1996; Schooler, Fiore, & Brandimonte, 1997), the contradictory research must also be reviewed. A few studies have found support for a small relationship between description accuracy and identification accuracy. For example, Wells (1985) has shown that description accuracy correlated significantly with identification accuracy ($r$ = .27). Additionally, Sporer (1996) reported a significant relationship between descriptor quantity (number of descriptors) and identification accuracy. That is, eyewitnesses who made correct identifications used more descriptors ($M$ = 6.52) than eyewitnesses who made incorrect identifications ($M$ = 5.16). Chan and McDermott (2007) provided a theoretical rationale for why a relationship between description accuracy and identification accuracy might exist. They suggested that recall and recognition should be related to the extent that both tasks rely on some common process.

Contrary to the research described previously suggesting no link between description accuracy and identification accuracy (Flexser & Tulving, 1978; 1982; Tulving & Thomson, 1973, Schooler and Engstler-Schooler, 1990), the Sporer (1996) and Chan and McDermott

(2007) studies suggest that there may, in fact, be a (weak) relationship there. However, any support for the relationship between recall and recognition has only been weak and has not yet been replicated across studies. Based on this lack of support for the description-identification relationship and the strong support in favor of no relationship, I hypothesized that the quality and quantity of descriptors used to describe a criminal will not be related to the accuracy of a subsequent identification (*Hypothesis 4*). Based on Schooler and Engstler-Schooler (1990) and Miessner and Brigham (2001), I expect the relationship between description accuracy and identification accuracy to be negatively correlated.

**Do people who provide a description perform differently on a subsequent identification task than those who do not?** As mentioned in the previous section, verbal overshadowing is evident when recognition is poorer in the description than in the no-description condition (e.g., Schooler & Engstler-Schooler, 1990). A meta-analysis of studies using this paradigm has shown a reliable, though small, negative effect of verbalization on face recognition (Meissner & Brigham, 2001). Importantly, in these studies, verbalization is elicited from memory *postencoding* of the face, *entails a detailed description*, and *recognition* of the target is tested in a lineup procedure.

It was originally proposed that words hinder recognition when there is a mismatch between the nonverbal knowledge required for successful completion of the "task in hand" and the verbal knowledge associated with describing or naming a stimulus. This general principle was termed the *modality mismatch assumption* (Schooler & Engstler-Schooler, 1990; Schooler, Fiore, & Brandimonte, 1997) and draws upon the well-known transfer-appropriate processing framework, the premise of which is that performance on memory tests benefits more when encoding operations overlap maximally with the retrieval demands of a

particular test (e.g., Morris, Bransford, & Franks, 1977). Several different accounts have since been put forward to explain how a mismatch between verbalization and the nonverbal demands of the memory test may arise. I briefly consider these here as they apply to face recognition.

In the face recognition domain, it is generally accepted that a mismatch between verbal and nonverbal knowledge is evident because it is difficult to capture in words information about a face (e.g., Ellis, Shepherd, & Davies, 1980; Schooler et al., 1997). Descriptions tend to emphasize the individual features of the face. In contrast, the subtle information concerning the spacing and relationships between facial features cannot be adequately described. In the face recognition domain, three separate accounts have been proposed to explain for how this mismatch between verbal and nonverbal knowledge may arise.

First, a post-encoding description may influence a participant's response criterion during the recognition task. Clare and Lewandowsky (2004) showed that participants who had previously described the target face, compared with those providing no description, were more likely to say that the target was "not present" in a recognition lineup (i.e., they adopted a more conservative response bias). The reason for this is not clear, but it has been suggested that because participants find describing a face difficult, they tend to infer more generally that their memory for the face is poor, and so they become reluctant to choose a target from the lineup (Clare & Lewandowsky, 2004). This account, however, does not explain why describing a face can still interfere with recognition when participants are forced to choose someone from the lineup.

Second, describing a face postencoding may lead to an inaccurate or imprecise description of the contents of the original memory. This new verbally biased representation successfully competes with the original visual memory and is inappropriately relied on at test, an account known as *retrieval-based interference* (Schooler & Engstler-Schooler, 1990; see also Meissner, Brigham, & Kelley, 2001). In this case, verbal overshadowing should be apparent when the contents of the description do not sufficiently aid memory; that is, poor quality descriptions should be correlated with poor memory performance. Indeed, some studies have revealed that descriptions containing more incorrect details tend to be associated with less accurate recognition performance (e.g., Finger & Pezdek, 1999; Meissner et al., 2001; see also a meta-analysis by Meissner et al., 2008). However, such a correlation has not always been observed, and these situations prove problematic for a retrieval-based interference account (e.g., Brown & Lloyd-Jones, 2003; Fallshore & Schooler, 1995; Schooler & Engstler-Schooler, 1990).

Finally, describing a face postencoding may shift the participant from using nonverbal processing previously applied to encode the face, to verbal processing. Verbal processing is then inappropriately applied at retrieval and impairs performance: a *transfer-inappropriate processing shift account* (e.g., Schooler, 2002). This shift in processing style is proposed to be relatively general, and it successfully explains those instances in which the negative effects of postencoding descriptions have been found to extend beyond the particular face that is described (e.g., Brown & Lloyd-Jones, 2002, 2003; Dodson, Johnson, & Schooler, 1997; Lloyd-Jones & Brown, 2008). This transfer-inappropriate processing shift account has historically assumed that verbal descriptions and featural processing are closely related and may elicit synonymous processing (Chin & Schooler, 2008; Schooler, 2002).

Descriptions that emphasize facial features result in a shift to featural processing at the expense of global/configural processing, and it is the latter type of processing that is best suited to face recognition (Diamond & Carey, 1986). In support of this theory, it has been shown that postencoding descriptions more generally impair performance on tasks that predominantly rely on global/configural knowledge, but not tasks that rely on featural/analytic knowledge.

In summary, the findings concerning postencoding descriptions have led to the assumption that nonverbal tasks negatively affected by verbalization are those that require global/configural knowledge for successful performance. This assumption is likely accurate because a description emphasizes feature-based information at the expense of global/configural information that is more difficult to describe. This diffuculity may lead to impaired performance because (a) the participant inappropriately relies on an imprecise or inaccurate verbal representation or (b) because the act of describing encourages a shift to a featural processing style that reduces the participant's ability to adequately access global/configural information (e.g., Chin & Schooler, 2008; Schooler, 2002). Based on this research, I expected to find similar results. That is, I expected to find impaired performance on the identification task for participants who engage in the description conditions compared to participants who are in the control condition (*Hypothesis 5a*).

**Does the timing of the description task affect identification accuracy?** There is essentially no empirical research regarding how the time after which eyewitnesses are asked to recall a criminal's appearance after viewing a crime (i.e., immediately or delayed) might influence that identification. This study was designed to fill that gap in the literature. The little we do know about the time at which eyewitnesses provide descriptions of criminals

after witnessing crimes comes from archival data. One such archival study of adult felony cases accepted by the San Diego Court District Attorney's Office found that the majority (58%) of descriptions given by eyewitnesses and victims of a crime were typically given within hours after the incident (Konečni & Ebbesen, 1992). More recent statistics are unavailable.

A few studies provide indirect evidence about whether the timing of a description task influences identification accuracy. For example, Wixted & Ebbesen (1997) had participants view a simulated crime and asked the participants to write (recall) details of the crime before picking the criminal from a lineup one to five days later. They found that the time between the crime and participants' recall of the crime (approximately 1-5 days) did not affect the accuracy of their identification.

Finger and Pezdek (1999) conducted three experiments in their study looking at many variables, including descriptions and identifications. In Experiment 3, Finger and Pezdek manipulated the time in which participants were asked to recall after viewing a face. After participants viewed a face, they were randomly assigned to 1 of 3 conditions: verbally describe the criminal immediately, after a 24 minute delay, or not all (control condition). A face identification recognition task followed immediately. Participants were significantly less accurate at identification in the immediate verbal description condition than those in the no description control condition. More importantly, however, participants in the delayed description condition performed equivalently to those in the no description control condition. These results suggest that the interference from the verbal overshadowing effect is released when there is a delay in recall, significantly improving identification accuracy to the same level of accuracy when no description is asked of the participant at all. Although some

researchers have observed similar effects (Yu and Geiselman, 1993), others have been unable to replicate these results (Schooler and Engstler-Schooler, 1990). If the time after which people are asked to describe a criminal does, indeed, influence accurate eyewitness identification, it is unclear what the effect would look like. Because the research is so mixed, specific predictions were not made about the effect of time of recall might have on recognition, but any effect will be investigated (*Hypothesis 5b*).

**Study Overview**

In this study, I used a conventional eyewitness paradigm. Participants viewed a brief breaking and entering scene, provided a description of the criminal either immediately, delayed (1-4 days later), or not at all (control group), and attempted to identify the criminal in a 6-person simultaneous lineup within 1-4 days. My hypotheses are reiterated below.

**Eyewitness Descriptions (Recall)**

**Hypothesis 1a.** I hypothesized that female participants would outperform male participants in the description accuracy and descriptor quantity of the criminal's description (female-superiority effect).

**Hypothesis 1b.** An own-gender bias for eyewitness descriptions was not hypothesized in this study, but was explored.

**Hypothesis 2.** I hypothesized that participants who were asked to recall immediately would have more accurate descriptions of the criminal compared to participants who were asked to recall after a delay or not at all (control condition).

**Eyewitness Identifications (Recognition)**

**Hypothesis 3a.** I hypothesized that females would outperform males in identification accuracy of the criminal (female-superiority).

**Hypothesis 3b.** Own-gender bias in identification was explored.

**Hypothesis 4.** I hypothesized that the quality and quantity of the eyewitness' description of a criminal would not be related to the accuracy of a subsequent identification.

**Hypothesis 5a.** I hypothesized that participants who engaged in the description conditions would have impaired performance on the identification task compared to participants who were in the control condition.

**Hypothesis 5b.** The effect of time of recall on identification accuracy was examined.

# Method

## Design

The study was a 2 (participant gender: male, female) × 2 (criminal gender: male, female) × 3 (timing of criminal description: immediate description, delayed description 1-4 days later, or no description) between-subjects factorial design. Participants were randomly assigned to conditions. The primary dependent variables were participants' photo identification accuracy, description accuracy, and descriptor quantity. Participants' pre- and post-confidence in the identifications and pre- and post-clarity of memory that they had for the criminal were also measured.

## Participants

Participants for this study were a sample of 543 U.S. adults (41.3% male, 58.7% female, 76.9% Caucasian, 8.6% African American, 14.5% Other, $M_{age}$ = 33.72). The study was administered using Qualtrics online surveying software. Participants volunteered through Amazon's Mechanical-Turk surveying system. The study was conducted in two parts. Participants were compensated $0.20 for participating in Part 1 and received an additional $0.50 if they participated in Part 2. One thousand eight hundred sixty-two participants completed Part 1. Of the 1,862 participants that completed Part 1, 627 continued to Part 2. Sixty participants were excluded because they failed to complete the study. Another 24 participants were excluded if they reported a distraction rating of greater than 10 (i.e., *extremely distracted*) in either part of the study.

The Institutional Review Board (IRB) approved this study on September 26, 2013, with an expiration date of September 25, 2014 (see Appendix A). All participants provided

informed consent before beginning each part of the study. Copies of the consent forms for Part 1 and Part 2 are provided in Appendix B.

**Materials**

**Crime video.** Two 40 second videos were recorded of either a male or a female criminal breaking into and burglarizing a home. The videos were filmed in black and white. Both actors were White and matched each others' movements. In the video, the criminal was shown entering a glass-paneled door, checking the surroundings, opening cabinet doors, placing stolen items into a bag, and briefly gazing toward the camera (ostensibly a surveillance system) prior to leaving. Screen shots of the crime videos are shown in the appendix (see Appendix C).

**Anagram filler task.** As in real life, eyewitnesses who view a crime may be asked to describe details of the crime within minutes after the initial viewing. What distracts eyewitnesses between the time of crime and recall in real-world scenarios is not entirely known, but we can speculate that eyewitnesses are not asked to describe the crime immediately, within seconds, after the crime. At best, eyewitnesses are approached by law enforcement for descriptive details within 15-20 minutes of a crime. For this reason, we chose to include an approximately 15 minute anagram filler task in order to simulate a situation as distracting as in a real-life crime scenario.

Each anagram was presented sequentially for 15 seconds and a prompt was provided for participants to type a solution. After the 15 seconds ended, participants were shown the next anagram. Before solving the 25 anagrams, participants were given 3 practice anagrams. All anagrams were randomly presented across participants.

The twenty-five 5-letter anagrams were chosen based on criteria established in Gilhooly and Johnson's (1978) study of anagram difficulty (see Appendix D). In their study, Gilhooly and Johnson created a list of 5-letter anagrams categorized via 10 measures of difficulty. Of those 10 measures, I selected words based on measures #3 and #5. Measure #3 was the *degree of letter order similarity* score, defined as the sum of the number of letters in the correct sequence in the anagram. For example, "ONGYA" has a letter order similarity score of 2 for its solutions "AGONY" ("ON" is the only letter sequence in correct order) while "ITRUF" (solution is "FRUIT") has a score of 4 since "IT" (2) and "RU" (2) are in the correct order. Gilhooly and Johnson's pilot studies indicated that this measure was a useful index of anagram similarity and that as this measure increased, so did solution rates. As a result, I chose 3 practice anagrams with scores ranging from 1-2. The 25 words used during the anagram task had anagram similarity scores ranging from 2-3. Measure #5 was the *GTZERO* score. GTZERO scores are calculated from the bigram frequency matrix and is the total number of bigrams with a frequency of greater than zero in the bigram frequency matrix[1]. For example, for the anagram "IGTHL" ("LIGHT") "HG," "HT," "HL," "GT," "TG," "TL," "LH," "LG," "LT" would all have a frequency of 0 in the first position. "IGTHL" would have a GTZERO score of 33. The more non-zero entries there are (the higher the GTZERO score), the greater the possible competing solutions that make the anagram harder to solve (Gihlooly & Johnson, 1978). The 3 practice anagrams in my study had GTZERO scores ranging from 28-63. The 25 words used during the anagram task ranged

[1] The bigram frequency matrix is constructed by drawing a table with 20 rows representing the possible bigrams (two letter sequences) and four column representing the four bigram positions in a five-letter word. The bigram rank is the number of entries in the table which have higher frequencies than the four correct entries (i.e., real bigram positions).

from 16-51. For more about the selection criteria of the anagrams see Gilhooly and Johnson's (1978) article on anagram difficulty (pp.66-68).

**Demographic information.** Participants were prompted to provide demographic information, including age, race, and gender (see Appendix E).

**Pre/Post confidence ratings.** Confidence was measured three times. Pre-confidence in ability to identify the criminal from a lineup was assessed twice, in both Part 1 and Part 2. Post-confidence in ability to identify the criminal from a lineup in the subsequent identification of the criminal was assessed once in Part 2. Confidence was assessed using an 11-point scale (0 = *not at all certain*; 10 = *absolutely certain*; see Appendix F). While pre- and post-confidence were measured, they were not analyzed for this thesis.

**Pre/Post clarity of memory rating.** Clarity of memory was measured three times. Pre-clarity of memory for the criminal in the video before identifying him/her in a lineup was assessed twice, in both Part 1 and Part 2. Post-clarity of memory in the subsequent selection of the perpetrator was assessed once in Part 2. Pre- and Post-clarity of memory were assessed using an 11-point scale (0 = *not at all clear*; 10 = *absolutely clear*; see Appendix F). While pre- and post-clarity of memory were measured, they were not analyzed for this thesis.

**Distractibility rating.** Because participants engage in the study on computers in an unknown environment, this study was not able to control for stimuli that might distract participants engaging in the study. As such, it is important to note that this measure of distraction is a self-report measure that is dependent upon participants' willingness to report whether they were truly distracted or not distracted while engaging in the study. This measure was included to ensure that participants paid a reasonable amount of attention to the crime video so as to make a reasonable identification of the criminal from the lineup.

Distraction was measured two times, at the end of both Part 1 and Part 2. Distraction during the study was assessed using an 11-point scale (0 = *not at all distracted*; 10 = *extremely distracted*; see Appendix F). This measure was used to exclude participants.

**Description task.** The computer monitor cued participants to, "Please provide a description of the criminal that you saw in the video at the beginning of the study. You should include a description of the criminal's physical appearance (examples include height, weight, hair style, etc.)" See Appendix G.

A pilot study was performed to obtain "accurate" descriptions of the male and female criminal targets. Descriptions of the target persons were obtained from 31 college-student raters (all 31 described both the male and female criminal) who viewed them for as long as they wished while completing their descriptions. Pilot subjects' responses were tabulated into means and standard deviation for the age, height, and weight variables. Modes were computed for the discrete variables (e.g., clothing, hair length, etc.).

In order to assess description accuracy, a checklist of the type frequently used by police was created. This checklist included each of the criminal's physical features. The checklist included 14 categories: race, sex, age, height, weight, body build, hair color, hair length jacket, shirt, pants, shoes, bag, and lock pick. In order to assess internal consistency, 30 (approximately 10%) of the descriptions were scored by two research assistants. Internal consistency reflects the degree to which judgments about a criminal description are consistent across observations or items thought to reflect the same descriptive dimension. An inter-rater reliability analysis using the Kappa statistic was performed to determine consistency among the two raters. The interrater reliabilities for the measures of description

accuracy ranged from .76 to 1.00. According to Landis and Koch (1977), these Kappa values reflect a substantial measure of agreement between the two raters.

As no precedent has been set for the scoring of a person's physical descriptions, I employed Pigott and Brigham's (1985) four measures of description accuracy. For description accuracy measure #1, responses for each description were divided into categories and scored as follows: 4 = *very accurate* (matching raters' modal response); 3 = *slightly accurate* (a small difference from raters' modal response); 2 = *slightly inaccurate* (a moderate difference from raters' modal response); and 1 = *very inaccurate* (a large difference from raters' modal response). Description involving the actual measured characteristics of the criminals (age, height, and weight) were scored in terms of their degree of standard deviation from the correct response. Factors not mentioned by participants were not included in the calculation of his or her mean description accuracy score. Table 1 depicts the 14 categories for the male and female criminals. Description accuracy measure #2 was derived by multiplying a subject's mean accuracy score by the number of traits that he or she attempted to describe. Description accuracy measure #3 was similar to the first except that the third measure treated participants' non-responses as zeros, such that an individual's mean score would be lowered by omitting any feature description. The last description measure, description accuracy measure #4, treated response as either right or wrong. For a subject's response to be correct, it had to match the raters' mean or modal response for that particular characteristic exactly. Correct responses were scored +1. All other responses were considered incorrect and scored as 0. Descriptor quantity, the number of descriptors provided by participants, was also recorded. The directions for scoring the description accuracy measures and descriptor quantity measure are in Table 2.

**Photo lineup.** A black and white six-photo lineup was comprised of the photo of the perpetrator and five foils who matched the verbal description of the perpetrator. This description was provided by a sample of 20 pilot participants (all 20 rated both the male and female criminal). Those students also viewed the initial lineup and provided feedback relative to any particular features of the persons or differences in backgrounds that could potentially influence identifications. That feedback was used to minimize bias in the lineups. The photos were presented simultaneously, rather than sequentially, since the majority of police use simultaneous lineups in eyewitness identification procedures (Wells & Loftus, 1984; Wells & Seelau, 1995). The photo of the perpetrator was randomly placed in positions 1 to 6 in the lineup. A checkbox was provided beneath each picture for participants to indicate which person they perceived to be the perpetrator. There was also a checkbox to indicate that the perpetrator was not present. The lineup is a 2 × 3 display, with three photos on the top row and three photos on the bottom row. Photo lineups for both the male and female criminal are included in Appendix H.

## Procedure

**Part 1.** Upon logging in to Amazon's Mechanical-Turk, participants were provided with an electronic copy of the Informed Consent form. If they agreed to offer consent, participants next watched the 40 second video-clip showing either a male or female burglarizing a home. Immediately following the video, all participants were next asked to complete the demographic questionnaire. Once those were completed, participants completed the filler anagram task. Male and female participants in the immediate recall condition were asked to provide a description of the criminal they saw in the video at the beginning of the study. The other two conditions (delayed recall group and control group) were not asked to

provide a description of the perpetrator during this first part of the research. All participants were next asked to complete three 11-point rating scales in order to a) indicate their confidence in their ability to identify the criminal from a photo lineup in the future, b) to rate the clarity of their memory they had for the criminal, and c) to rate the extent to which they were distracted while participating in the study (Bradfield, Wells, 2000; Wells, Small, & Penrod, 1998). Participants were then thanked for their time and commitment to the study and were paid $0.20 for their participation in Part 1. They were reminded to wait 1-4 days in order to participate in Part 2 of the study.

**Part 2.** Only participants from Part 1 participated in Part 2 of the study. After completing the informed consent form, participants were reminded of the video they viewed in Part 1. The delayed recall group who were not asked to complete a description in Part 1 were asked to provide a description of the criminal in the video in Part 1. The control group were not. Finally, as in a real-life simultaneous lineup procedure, participants in each condition were told, "You will now see a photo lineup that may contain the person who committed the burglary. Naturally, the perpetrator may not appear in the lineup at all. Your task is to choose the number that is shown below the picture of the person you feel committed the crime or choose the number indicating the perpetrator is not included in the lineup" (Steblay, 1997; Wells, 1978). Participants were shown a male or female lineup based upon the gender of the criminal they viewed in the video in Part 1. Following the lineup task, participants were asked to use the 11-point scales provided to estimate their confidence that they made the correct selection decision, the clarity of their memory of the video and the criminal, and the extent to which they may have been distracted while completing this study (Bradfield, Wells, 2000; Wells et al., 1998). At the conclusion of the experiment, participants

were thanked for their time and commitment to the study and were paid $0.50 for their participation in Part 2.

## Results

### Eyewitness Descriptions (Recall)

**Preliminary analysis.** Based on Pigott and Brigham's (1985) free recall scoring method, four description accuracy scores and a descriptor quantity score were computed for all participants who were in the recall conditions ($N$ = 294). First, the measures of description accuracy and measure of descriptor quantity were correlated (Table 3). As would be expected, each of the four measures of description accuracy were significantly correlated with all of the others, $p < .001$. The correlations ranged from .33 - .80. In addition, three of the description accuracy measures were found to be significantly correlated with the measure of descriptor quantity, $p < .001$. The correlations ranged from .43 - .94. The only non-significant correlation was between description accuracy measure #1 and descriptor quantity, $r(924) = -.03$.

Before conducting planned comparisons based on my hypotheses for description accuracy and descriptor quantity, I first report the results of the full model that included main effects for participant gender, criminal gender, recall time, and their interactions. A 2 (participant gender) × 2 (criminal gender) × 2 (time of recall) factorial ANOVA was performed on each of the four description accuracy measures and the descriptor quantity measure. The omnibus test was significant for description accuracy measure #1, $F(7, 286) = 3.60$, $p = .001$, $R^2$=.08. There was a significant main effect of criminal gender on description accuracy measure #1, $F(7, 286) = 15.23$, $p < .001$, $R^2 = .08$. Omnibus tests for description accuracy measures #2, $F(7, 286) = 1.60$, $p = .134$, $R^2 = .04$, description accuracy measures

#3, $F(7, 286) = 1.33$, $p = .234$, $R^2 = .03$, description accuracy measure #4, $F(7, 286) = .54$, $p = .801$, $R^2 = .01$, and descriptor quantity measure, $F(7, 286) = 1.80$, $p = .087$, $R^2 = .04$ were not significant. No main effects or interactions were statistically significant for those measures.

While the full factorial ANOVAs did not support my hypotheses, those analyses demanded more power, which might have reduced the ability to detect specific effects. Therefore, I chose to also test specific hypotheses with single factor ANOVAs.

**Gender and eyewitness descriptions.** In order to test *Hypothesis 1a*, that female participants would on average have higher scores than male participants in criminal description accuracy and quantity (female-superiority main effect), a single factor ANOVA with participant gender as the factor was done separately on each of the four measures of description accuracy and descriptor quantity. Gender of participant was related to description accuracy measure #2 $F(1, 292) = 8.92$, $p = .003$, $R^2 = .03$), description accuracy measure #3, $F(1, 292) = 4.43$, $p = .036$, $R^2 = .02$, and descriptor quantity measure, $F(1, 292) = 8.46$, $p = .004$, $R^2 = .03$. Female participants had higher average scores than male participants on each of these measures (Table 4). However, gender of participant was not related to description accuracy measure #1, $F(1, 292) = .67$, $p = .412$, $R^2 = .002$, nor description accuracy #4 $F(1, 292) = .63$, $p = .429$, $R^2 = .002$. These results show partial support for the hypothesis for a female-superiority main effect for description accuracy and full support for the female-superiority main effect for descriptor quantity.

In order to further examine *Hypothesis 1b*, an own-gender bias, the interaction between gender of participant and gender of criminal was examined. Separate 2 (gender of participant) × 2 (gender of criminal) ANOVAs were done for each measure of description accuracy and descriptor quantity. A significant participant gender by criminal gender

interaction was found for description accuracy measure #1, $F(1, 292) = 14.56$, $p < .001$, $R^2 = .05$. Subsequent analyses showed that female participants' description accuracy measure #1 scores were significantly higher, $F(1, 102) = 2.17$, $p < .001$, for the female criminal ($M = 3.38$, $SD = .36$) than for the male criminal ($M = 3.17$, $SD = .40$). However, there was no difference, $F(1, 102) = 3.21$, $p = .07$, in male participants' description accuracy measure #1 scores for the female ($M = 3.29$, $SD = .36$) or male criminal ($M = 3.16$, $SD = .36$) (Figure 2). Interactions for description accuracy measure #2, $F(1, 292) = 3.63$, $p = .058$, $R^2 = .01$, description accuracy measure #3, $F(1, 292) = 1.65$, $p = .200$, $R^2 = .01$, description accuracy measure #4, $F(1, 292) = .23$, $p = .634$, $R^2 = .001$, and descriptor quantity measure, $F(1, 292) = .35$, $p = .554$, $R^2 = .001$, were not statistically significant. These results show little support for a participant gender by criminal gender interaction (own-gender bias) for description accuracy and descriptor quantity.

**Effects of time of recall on description accuracy**. To test *Hypothesis 2*, that participants who were asked to recall immediately would have high description accuracy scores and descriptor quantity score of the criminal compared to participants who were asked to recall after a delay, I performed five separate *F*-tests for each description accuracy measure and for descriptor quantity.

No statistically significant differences between the immediate and delayed recall groups in mean scores were found for any of the four description accuracy measures nor descriptor quantity. These results do not support the hypothesis that participants who were asked to recall earlier would have more accurate descriptions and descriptor quantity compared to the participants who were asked to recall after a delay.

**Eyewitness Identifications (Recognition)**

**Preliminary Analysis.** As in the case of the description accuracy and descriptor quantity analyses, before planned comparisons for identification accuracy based on my hypotheses, I first report the results of the full logistic regression model that included main effects for participant gender, criminal gender, recall time, and their interactions. I first examined the proportion of correct identifications (hits) made by the participants in the experiment. A 2 (participant gender) × 2 (criminal gender) × 3 (time of recall) binary logistic regression analysis was conducted for hits. The overall rate of prediction accuracy for hits for the null model was 57.50%. The overall rate of prediction accuracy for hit rates for the full model was 63.0%, $\chi^2(11) = 44.42$, $p < .001$. However, the only statistically significant effect was the main effect of criminal sex, Wald $\chi^2(1) = 9.22$, $p = .002$, Exp(B) = 3.16; therefore, the odds of a participant identifying the criminal was approximately 3 times higher for the male criminal than for the female criminal.

**Gender and eyewitness identifications.** In order to test *Hypothesis 3a*, that female participants would have more accurate identifications of the criminal compared to male participants (female-superiority main effect), a binary logistic regression analysis consisting of participant gender was done for hits. The overall rate of prediction accuracy for hits for the null model was 57.50%. The overall rate of prediction accuracy for hits for the full model did not increase, and was also 57.50%. Gender of participant was not related to hit rates, $\chi^2(1) = .87$, $p = .352$. These results do not support the hypothesis that female participants would have more accurate identifications of the criminal compared to male participants (female-superiority main effect).

In order to further examine *Hypothesis 3b*, for a participant gender by criminal gender interaction (own-gender bias), a 2 (participant gender) × 2 (criminal gender) binary logistic regression analysis was done for hits. The overall rate of prediction accuracy for hits for the null model was 57.50%. The overall rate of prediction accuracy for hits for the full model was 62.60%, $\chi^2(1) = 11.29$, $p = .001$. A significant participant gender by criminal gender interaction was not found for hits, Wald $\chi^2(1) = .03$, $p = .862$, Exp(B) = .93. This result is consistent with the results of the full model examined in the preliminary analysis. When the participant gender by criminal gender interaction was examined in the full model, consisting of participant gender, criminal gender, and time of recall, the interaction was also not statistically significant. These results show no support the hypothesis for a participant gender by criminal gender interaction (own-gender bias) in eyewitness identifications.

**Description-identification accuracy relationship.** In order to test *Hypothesis 4*, that description accuracy and descriptor quantity would not be related to identification accuracy, all four measures of description accuracy and measure of descriptor quantity were correlated with hits. Results were consistent with this hypothesis. None of the measures of description accuracy nor descriptor quantity were significantly correlated with hits.

**Effects of recall versus no recall on identification accuracy.** In order to test *Hypothesis 5a*, that participants who engaged in the description conditions would have impaired performance on the identification task compared to the control condition, a binary logistic regression analysis consisting of the recall groups and no recall (control) group was done for hits. The overall rate of prediction accuracy for hit rates for the null model was 57.50%. The overall rate of prediction accuracy for miss rates for the full model did not

increase, and was also 57.50%. The act of recalling the criminal was not related to hit rates, $\chi^2(1) = .11$, $p = .737$, failing to support the hypothesis.

**Effects of time of recall on identification accuracy.** In order to test *Hypothesis 5b*, any effects of time of recall on identification accuracy, three separate binary logistic regression analyses were done for hits. The effect of recalling immediately versus recalling after a 1-4 day delay was related to hits, $\chi^2(1) = 5.07$, $p = .021$, Exp(B) = .58. The delayed recall group had a higher percentage of hit rates (64.3%) than the immediate recall group (51.2%). The effect of recalling after a 1-4 day delay versus not recalling at all was not related to hits, $\chi^2(1) = 1.29$, $p = .262$, Exp(B) = 1.30. The effect of recalling immediately versus not recalling at all was also not related to hits, $\chi^2(1) = 2.01$, $p = .164$, Exp(B) = .75.

## Discussion

I examined various system and estimator variables that might affect description accuracy and descriptor quantity for eyewitnesses' recall of a criminal and their accuracy of recognition of a criminal from a lineup. The results of my study showed limited support for gender differences in description accuracy and identification accuracy. However, I found that women provided longer descriptions (descriptor quantity) compared to men. My study also found that the time after which participants were asked to recall a criminal's description did not affect description accuracy nor descriptor quantity. However, when examining how time of recall might affect identification accuracy, results revealed that participants who were asked to recall after a delay had greater identification accuracy than participants who recalled immediately. And as expected, this study revealed no relationship between description accuracy and identification accuracy. These findings are discussed in more detail below.

**Eyewitness Descriptions (Recall)**

**Gender and eyewitness descriptions.** Relative to *Hypotheses 1a and 1b*, I examined the relationship between gender and eyewitness descriptions. I hypothesized that female participants would be superior to male participants in description accuracy and that they would use more descriptors to describe the criminal (female-superiority main effect). Interactions between participant gender and criminal gender (own-gender bias) were also examined.

***Descriptor quantity*.** The results showed support for the hypothesis of a female-superiority main effect for descriptor quantity. On average, female participants provided more descriptors than their male counter parts. This finding is consistent with other research that has found that women provided more descriptors when asked to describe people (i.e., Areh, 2011) and events (i.e., Lindoln & Christianson, 1998). When participant gender and criminal gender were examined, support for an interaction (own-gender bias) for descriptor quantity was not found.

***Description accuracy*.** The results showed no convincing support for the female-superiority main effect relative to the accuracy of the description of the criminal. On average, female participants had higher scores than male participants on description accuracy measures #2 and #3. However, there were no differences between male and female participants' scores on description accuracy measures #1 and #4. Very little support for a participant gender by criminal gender interaction (own-gender bias) for description accuracy was found. However, for description accuracy measure #1, female participants' scores were significantly higher when recalling the female criminal than when recalling the male criminal. No differences were obtained between male participants' scores for recalling either

the female or male criminals. No interactions for description accuracy measure #2, #3, or #4 were found.

The data provided no convincing support for a female superiority main effect or for an own-gender bias in description accuracy. Only two out of the four measures of description accuracy showed support for a female-superiority main effect, and only one out of four measures showed support for an own-gender bias. Why was support found for some measures of description accuracy and not others? This could be explained by the observed effect sizes for the description accuracy measures. Even when effects were detected, such as for description accuracy measures #1, #2, and #3, effect sizes associated with these measures were small ($R^2$ ranged between .03-.08). Because the sample size ($N$ = 294) was not small, the power to detect effects, if they existed, would not have been a problem.

Another possible explanation for the inconclusive findings of my study could be due to how I chose to score description accuracy. Unlike my study, earlier studies that have found support for a female-superiority main effect (i.e., Areh, 2011; Lindholn & Christianson, 1998) and own-gender bias (i.e., Powers et al., 1979), measured recall using checklists. For instance, Areh (2011) had participants witness a crime after which they were asked to answer a 20-item list of questions regarding specific characteristics of the criminal in the video. Items included height, body-build, gender, ethnicity, etc. In other words, Areh (2011) used a cued recall method while I tried to assess memory using free recall, which might have produced different results.

Overall, it is unclear why only some measures of description accuracy show support for a female-superiority main effect and own-gender bias. It might just be difficult to obtain a reliable measure of description accuracy when scoring free recall material. It should be

recognized that a standardized measure of description accuracy has yet to be developed for free recall tasks. Therefore, I suggest that future research should be done to establish a more valid measure of description accuracy for recall tasks.

**Effects of time of recall on description accuracy.** Relative to *Hypothesis 2*, I examined the relationship between the timing of the recall of the criminal and subsequent description accuracy. My hypothesis that participants who were asked to recall the criminal immediately would demonstrate greater accuracy than participants who were asked to recall the criminal after a delay was not supported. These results are inconsistent with the forgetting curve of Ebbinghaus that suggests that recall is more accurate soon after the experience and gradually decreases as time passes (Ebbesen & Rienick, 1998; Kassin et al., 2001, Laney & Loftus, 2009). However, the results are commensurate with those of Wixted and Ebbesen (1997) that time between a crime and participants' recall of the crime did not affect description accuracy.

One possible reason for the finding of no relationship between time of recall and description accuracy could be due to the time length after which participants in the delayed recall group were asked to recall the criminal. Wixted and Ebbesen (1997) had participants recall details of a criminal 1-5 days after viewing the crime. My participants had a similar delay: 1-4 days. However, most studies that examined the relationship between recall and description accuracy did so over a period of weeks and months. For these studies, a statistically significant decrease in description accuracy was observed (Ebbesen & Rienick, 1998; Laney & Loftus, 2009). My results combined with these earlier findings, suggest that in order for a decrease in description accuracy to be seen, the time between a viewing of a crime and recall may need to exceed 4-5 days.

Another possible explanation for my null results, might be due to *retroactive interference*, which occurs when newly acquired information inhibits ability to recall previously acquired information. It is possible that my participants engaged in behaviors after viewing the crime video that interfered with their ability to recall the crime. In a real-world criminal investigation, such retroactive interference could be a result of various estimator and system variables. For instance, eyewitnesses' ability to recall a crime might be distorted due to media coverage of a crime. The information presented during popular media coverage of a crime could potentially interfere with an eyewitness' true recall of the crime. My results are supported by the retroactive interference hypothesis, and suggest that what happens between the time an eyewitness views a crime and the time they are asked to recall it, should be considered. This lapse of time that should be considered in any investigation dependent upon an eyewitness' testimony of that relative to the crime, as their memory for the crime could be influenced by variables that may interfere with an eyewitness' ability to accurately recall.

**Eyewitness Identifications (Recognition)**

**Gender and eyewitness identifications.** Relative to *Hypotheses 3a and 3b*, I explored the accuracy of eyewitness identifications as a function of participant and criminal gender. I hypothesized that female participants would provide more accurate identifications (hits) of the criminals than male participants (female-superiority main effect). The interaction between participant gender and criminal gender (own-gender bias) was also examined. The results do not support the hypothesis that female participants would provide more accurate identifications of the criminal compared to male participants. When testing for an interaction between participant gender and criminal gender (own-gender bias) on identification accuracy, a significant interaction was not found. Furthermore, when the interaction was

examined in the full model consisting of all three factors (participant gender, criminal gender, and time of recall), no statistically significant interactions were observed. These results do not support past research that demonstrated own-gender bias interactions (i.e., Areh, 2011; Lewin & Herlitz, 2002; Rehnman & Herlitz, 2007).

One reason why the results in my study might be inconsistent with previous research that support a female-superiority main effect in identification accuracy could be due to the criminal gender main effect found for identification accuracy. That is, the male criminal was much easier to identify by participants compared to the female criminal. My results revealed that the odds of a participant identifying the criminal was approximately three times greater for the male criminal than for the female criminal. This difference occurred despite efforts to create fair and equivalent lineups for the male and female criminal based on feedback provided by 20 pilot participants. The large effect size associated with criminal gender seems to have accounted for a large proportion of the total hit variance in the present study. Of the total amount of variability, criminal gender accounted for so much of the total variance that little variability remained for the other factors. In addition, the large amount of residual variance in the study seems to have masked any effects that might have otherwise existed. In the future, careful pretesting of target faces should be employed.

**Recall-identification relationships.** Relative to *Hypothesis 4*, I examined the description accuracy-identification accuracy relationship and the descriptor quantity-identification accuracy relationship (Flexser & Tulving, 1978; Goldstein at al., 1979; Howells, 1938; Tulving & Thomson, 1973; Wolfskiel & Brigham, 1985). Based upon the research that found no relationship between identification accuracy and either description accuracy or descriptor quantity, I hypothesized that there would be no relationship between

them. Although verbal descriptions are an important part of police work, the results from my study provide no support for the validity of the Supreme Court's guideline that suggests that the accuracy of an eyewitness' prior description of a criminal should be considered regarding subsequent eyewitness identifications. Contrary to the Supreme Court's guideline, the results of my study suggest that it cannot be assumed that eyewitnesses who are accurate in describing a criminal will also be accurate at identifying that criminal from a lineup.

**Recall versus no recall.** Relative to *Hypothesis 5a*, I examined the relationship between the act of recall and recognition. Specifically, I hypothesized that participants who were asked to describe the criminal would have impaired performance on the identification task than participants who did not describe the criminal verbally—a demonstration of verbal overshadowing. Although many studies have replicated the verbal overshadowing effect (i.e., Dodson at al., 1997, Fallshore & Schooler, 1995; Finger & Pezdek, 1999), failed attempts at replicating the effect also exist (i.e., Meissner et al., 2001; Yu & Geiselman, 1993). My study also failed to find a verbal overshadowing effect.

A practical concern is that verbal overshadowing should have an important influence on the manner in which law-enforcement officials obtain information from eyewitnesses. For instance, law enforcement officers often ask eyewitnesses for physical descriptors of criminals. Furthermore, law enforcement might depend on the accuracy of an eyewitness' recall in order to develop a sketch of a criminal and/or develop a lineup for subsequent identification. If the identification of a criminal is impaired following a verbal description, then it would seem important to inform law-enforcement officials of the potential harm in such procedures. However, because my attempt to replicate verbal overshadowing was unsuccessful (like so many other researchers), the suggestions that can be made to law-

enforcement officials at this time regarding recall versus no recall affecting recognition are unsatisfying. Further research assessing the effects of recall on recognition, like that of describing a criminal and later identifying the criminal in a lineup, needs to be done to clarify the relationship between recall and subsequent identification.

**Timing of recall affecting identification accuracy.** While no differences in identification accuracy were found between the recall groups and the no recall group (*Hypothesis 5a*), relative to *Hypothesis 5b,* I sought to examine whether the time in which participants were asked to recall would affect identification accuracy. In my study, some participants were asked to recall the criminal immediately after viewing the crime video (immediate recall group). Other participants were asked to recall after a 1-4 day delay (delayed group). Another group was not given the recall task at all (no recall group). The identification accuracy of the immediate recall group was inferior to that of the delay recall group. However, the no recall group hit rate did not significantly differ from either of the recall groups. Inserting a 1-4 day delay between the crime and recall task resulted in *verbal facilitation*; identification accuracy was significantly better in the delayed recall groups than in the immediate recall group. This finding is consistent with the results of Finger and Pezdek (1999) in which identification accuracy memory was best after a time delay was inserted between the crime event and time of recall. However, my finding is inconsistent with previous research that found time of recall was unrelated to subsequent recognition tasks (Tulving & Thomson, 1973). My finding that inserting a time delay between the crime and recall task improved facial identification accuracy suggests that verbally describing a face does not impair memory for the original face viewed. In fact, my results suggest that recalling a criminal's description after a delay might improve identification accuracy.

In addition to the theoretical contributions of my findings, research concerning time of recall on recognition has important real-world implications as well. For instance, often after viewing a crime, eyewitnesses are interviewed by a law-enforcement officer who requests details of the crime as well as a criminal's physical description (Sporer, 1996). In the real world, these interviews can be conducted either immediately or days, weeks, or months following the crime. The point in time after which eyewitnesses are asked to recall a criminal's description could influence their memory for the criminal, and thus, affect subsequent identification of the criminal. Asking eyewitnesses to engage in recall after some time has passed, at least 1-4 days, could have some utility for later identification. Perhaps, law-enforcement officials should consider waiting 1-4 days after a crime before asking eyewitnesses to provide a description of a criminal. According to my results, waiting to recall (as opposed to recalling immediately) could enhance an eyewitness' memory and improve identification accuracy in a lineup. This is contrary to the forgetting curve and retroactive interference theories previously discussed. More research needs to be done in order to examine why recalling later (versus immediately) might improve recognition. In addition to asking eyewitnesses to recall at a particular time following an event, another strategy that law-enforcement might consider, is to not ask eyewitnesses to recall at all. While our results revealed an improvement in identification accuracy after a 1-4 day delay, our results also suggest that there is no difference in identification accuracy between the no recall group and either of the recall groups. In other words, eyewitness' memories might be more accurate if law enforcement did not ask eyewitnesses to recall events or a criminal's description at all. Not asking eyewitnesses to recall criminals could eliminate specific investigative issues. For instance, it is possible that law enforcement might create a photo lineup based on inaccurate

recall from a criminal. Such a lineup consisting of suspects based on an erroneous description could result in an identification of an innocent suspect. On the other hand, while not asking for description at all might yield more accurate identifications, it should be recognized that such practices would be impractical. Many common police procedures during an investigation rely heavily on eyewitness' descriptions of criminals. For instance, law enforcement count on eyewitness' descriptions of criminals to help them compile photo lineups based on that description. Police and sketch artists also depend on eyewitness' descriptions of criminals when searching for the likely perpetrators of a crime. By not asking eyewitnesses to recall such information would deprive law enforcement of criminal descriptions needed for such procedures.

**From M-Turkers to Real-Life Eyewitnesses**

In my study, I used a United States sample obtained through Amazon's Mechanical-Turk (M-Turk). There are strengths and weakness of this approach. Previous research suggests that conducting studies on M-Turk can provide results that are similar to findings using conventional samples (i.e., psychology undergraduate students; Birnbaum, 2000; Nosek, 2007). As in previous studies using M-Turk samples, my participants were more likely to be younger, overeducated, and underemployed than a representative sample from the community. We can also assume that all participants had computer and internet access and chose to complete this study for financial compensation. In that sense, it does not provide a random sample of the population. Because M-Turkers select the studies in which to participate, choosing to engage in my study may imply that they differ systematically from other M-Turkers or citizens who choose to participate in other studies.

In addition to self-selection issues, using M-Turk samples threatens strict experimental control over participants. For example, because participants are participating online and not in a controlled lab environment, it was not possible for me to observe participants' behaviors while engaging in the experiment. However, several provisions were put in place to increase the liklihood that participants were attentive during the experiment. First, all participants were told to pay attention to their screens during the time of the experiment. Second, participants were told that their compensation for completing the study was dependent upon their monitored quality of participation. For example, should their performance on various measures of the study (i.e. demographic questionnaire, anagram task, recall task, etc.) appear to be done in haste or incompletely, they would not be compensated for their participation. Third, participants were asked to indicate how distracted they were during the experiment. As previously mentioned, the few participants who indicated that they were *extremely distracted* were eliminated from data analysis. Last, participants' response times to various questions and total time taken to complete the study were recorded. These data were used to identify participants who did not take the study seriously and/or rushed through the study as well as those who took an unreasonably long time to complete the study. Participants with either unusually short or long times were eliminated from data analysis. While several provisions were employed to ensure a controlled environment, it must be recognized that no matter how many precautions were taken, lack of control over participants could never be eliminated using an online experimental procedure. For instance, participants in my study viewed a crime in Part 1. It could be that they were likely aware that they would have to recall and/or recognize the criminal in Part 2.

While lack of experimental control might be a major limitation of my study, the integrity of the study could also be viewed as a strength. In a real-life crime, eyewitness' attention and distraction during a crime cannot be controlled. In fact, system and estimator variables vary according to each crime environment and each eyewitness. It could be argued that the lack of control in both real-world crimes and online simulated crimes might produce comparable eyewitness experiences. Thus, perhaps the lack of experimental control in the current experiment might not be viewed as a limitation in methodology, but rather be viewed as a potential strength in its ecological validity.

**Future Research and Reforms**

Eyewitness descriptions and identifications are two of the fundamental investigative practices in the criminal justice system. The police, prosecutors and the courts have long relied on the testimony of eyewitnesses and victims to identify, prosecute, and convict individuals (Innocence Project 2010; National Institute of Justice, 1999, 2003; Wells & Seelau, 1995).

However, research on eyewitness memory over the past three decades, as well as the increasing attention to wrongfully convicted individuals, has raised questions and concerns about eyewitness memory to the point that it can no longer be given the very high level of credibility that it received in the past (Innocence Project 2010). In fact, human memory research findings have shown that in many situations, especially those that are confrontational or traumatic, human memory can be faulty. Scientists in this field have sought to better understand how memory works and to apply that knowledge to real-world situations where human memory is critical.

DNA exoneration cases have exposed eyewitness error as the predominant factor in wrongful convictions (Innocence Project 2010). Over a decade ago, concern for the validity of eyewitness testimony inspired joint action among law enforcement, legal professionals, and researchers, resulting in the 1999 publication of *Eyewitness Evidence: A Guide for Law Enforcement* by the National Institute of Justice (NIJ). The NIJ Guide used the available body of scientific research and best practices in law enforcement to provide recommendations for effective eyewitness identification procedures.

As the body of scientific research in eyewitness memory continues to grow and additional field and research studies are completed, one can expect that states and local jurisdictions will continue to examine their eyewitness identification policies and procedures. Where state legislature and policy makers do not act, it is likely that these reforms will not continue through the work of law enforcement leaders, prosecutors, criminal defense attorneys, and other advocates. However, it is my hope that the legal system will consider research regarding eyewitness memory, and modify any questionable procedures with improved practices based on quality research.

For instance, my results suggest that gender does not seem to affect description accuracy or identification accuracy. However, female participants provided more descriptors of the criminals compared to male participants. Knowing that men and women differ in descriptor quantity adds insight to eyewitness description procedures. Specifically, law enforcement might expect descriptions provided by women to be longer and more detailed compared to men. I suggest that it is in the legal system's greater interest to be aware of any eyewitness procedures where men and women are asked to engage in descriptions, as this difference in descriptor quantity might be observed. Although there is no known

consequence of providing longer descriptions in real-world situations, you might expect that eyewitness who provide more descriptors of a crime and/or criminal appear more credible or knowledgeable. Law enforcement might be persuaded to trust an eyewitness who appears to be familiar with the crime at a detailed level. However, based on my study along with other research, descriptor quantity alone appears to be unrelated to identification accuracy. It should not be assumed that an eyewitness who is able to provide longer descriptions is more accurate.

Another aspect of eyewitness memory that legal systems might consider is the time after which eyewitnesses are asked to recall a criminal's description. According to my findings, delayed recall of a criminal might be better than immediate recall. Although, perhaps not practical relative to lineup development, not being asked to recall at all might also be better than recalling at any point in time. Because asking eyewitnesses to describe criminals is a common task in eyewitness procedures, it would be of utility to the legal system to be aware of how time of recall might affect memory accuracy.

It is my hope that more of an effort will be made by the legal system to understand eyewitness research. As research in eyewitness memory progresses, routine assessment of current practices will continue to be useful in informing researchers and policymakers on areas that require additional attention and reform. Research in the laboratory and the field must continue, so that researchers can more fully understand eyewitness memory. Once researchers understand the variables that contribute to eyewitness misidentifications, improvements to existing identification procedures can be made to assist in criminal investigations.

## References

Areh, I. (2011). Gender-related differences in eyewitness testimony. *Personality And Individual Differences*, *50*, 559-563. doi:10.1016/j.paid.2010.11.027

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law And Human Behavior*, *25*, 475-491. doi:10.1023/A:1012840831846

Birnbaum, M. H. (Ed.). (2000). Psychological experiments on the Internet. San Diego, CA: Academic Press.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality And Social Psychology Bulletin*, *15*, 19-25. doi:10.1177/0146167289151002

Boyce, M., Beaudry, J., & Lindsay, R. L. (2007). Belief of eyewitness identification evidence. In R. L. Lindsay, D. F. Ross, J. Read, M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 501-525). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Bradfield, A. L.,& Wells, G.L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law & Human Behavior*, *24*, 581-582.

Brigham, J. C., & Barkowitz, P. (1978). Do 'They all look alike'? The effect of race, sex, experience, and attitudes on the ability to recognize faces. *Journal Of Applied Social Psychology*, *8*, 306-318. doi:10.1111/j.1559-1816.1978.tb00786.x

Brigham, J. C., Bennett, L., Meissner, C. A., & Mitchell, T. L. (2007). The influence of race on eyewitness memory. In R. L. Lindsay, D. F. Ross, J. Read, M. P. Toglia (Eds.), *The handbook of eyewitness psychology, Vol II: Memory for people* (pp. 257-281). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Brigham, J. C., & WolfsKeil, M. P. (1983). Opinions of attorneys and law enforcement personnel on the accuracy of eyewitness identifications. *Law And Human Behavior, 7*, 337-349. doi:10.1007/BF01044736

Brown, C., & Lloyd-Jones, T.J. (2002). Verbal overshadowing in a multiple face presentation paradigm: Effects of description instruction. *Applied Cognitive Psychology, 16, 873-885.*

Brown, C., & Lloyd-Jones, T.J. (2003). Verbal overshadowing of multiple face and car recognition: Effects of within- versus across-category verbal descriptions. *Applied Cognitive Psychology, 17,* 183-201.

Brown, E., Deffenbacher, K., & Sturgill, W. (1977). Memory for faces and the circumstances of encounter. *Journal Of Applied Psychology, 62*, 311-318. doi:10.1037/0021-9010.62.3.311

Chan, J. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 33*, 431-437. doi:10.1037/0278-7393.33.2.431

Chance, J. E., & Goldstein, A. G. (1996). The other-race effect and eyewitness identification. In S. Sporer, R. S. Malpass, G. Koehnken (Eds.) , *Psychological issues in eyewitness identification* (pp. 153-176). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chance, J., Goldstein, A. G., & McBride, L. (1975). Differential experience and recognition memory for faces. *The Journal Of Social Psychology, 97*, 243-253. doi:10.1080/00224545.1975.9923344

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal Of Cognitive Psychology, 20*, 396-413. doi:10.1080/09541440701728623

Clare, J., & Lewandowsky, S. (2004). Verbalizing facial memory: Criterion effects in verbal overshadowing. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 30*(4), 739-755. doi:10.1037/0278-7393.30.4.739

Cross, J. F., Cross, J., & Daly, J. (1971). Sex, race, age, and beauty as factors in recognition of faces. *Perception & Psychophysics, 10*, 393–396.

Davies, G. (1992). Influencing public policy on eyewitnessing: Problems and possibilities. In F. Lösel, D. Bender, T. Bliesener (Eds.) , *Psychology and law: International perspectives* (pp. 265-274). Oxford, England: Walter De Gruyter. doi:10.1515/9783110879773.265

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal Of Experimental Psychology: General, 115*, 107-117. doi:10.1037/0096-3445.115.2.107

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition, 25*, 129-139. doi:10.3758/BF03201107

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal Of Applied Psychology, 83*, 745-762. doi:10.1037/0021-9010.83.5.745

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1980). The deterioration of verbal descriptions of faces over different delay intervals. *Journal of Police Science and Administration, 8,* 101– 106.

Fahsing, I. A., Ask, K., & Granhag, P. A. (2004). The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *Journal Of Applied Psychology, 89,* 722-729. doi:10.1037/0021-9010.89.4.722

Fallshore, M., & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal Of Experimental Psychology: Learning, Memory, And Cognition, 21,* 1608-1623. doi:10.1037/0278-7393.21.6.1608

Farrington, D. P., & Lambert, S. (1997). Predicting offender profiles from victim and witness descriptions. In J. L. Jackson, D. A. Bekerian (Eds.) , *Offender profiling: Theory, research and practice* (pp. 133-158). Hoboken, NJ: John Wiley & Sons Inc.

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal Of Applied Psychology, 84,* 340-348. doi:10.1037/0021-9010.84.3.340

Flexser, A. J., & Tulving, E. (1978). Retrieval independence in recognition and recall. *Psychological Review. 85,*153-171.

Flexser, A. J., & Tulving, E. (1982). Priming and recognition failure. *Journal of Verbal Learning and Verbal Behavior, 21,* 237-248.

Gilhooly, K. J., & Johnson, C. E. (1978). Effects of solution word attributes on anagram difficulty: A regression analysis. *Quarterly Journal of Experimental Psychology, 30,* 57-70.

Goldstein, A. G., Chance, J. E., & Schneller G. R. (1989). Frequency of eyewitness identification in criminal cases: A survey of prosecutors. *Psychological Bulletin, 21,* 71-74.

Goldstein, A. G., Johnson, K. S., & Chance, J. (1979). Does fluency of face description imply superior face recognition? *Bulletin of the Psychnomic Society, 13,* 15-18.

Herlitz, A., Nilsson, L., & Bäckman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25,* 801-811. doi:10.3758/BF03211324

Herlitz, A., & Rehnman, J. (2008). Sex differences in episodic memory. *Current Directions In Psychological Science, 17,* 52-56. doi:10.1111/j.1467-8721.2008.00547.x

Horry, R., Memon, A., Wright, D. B., & Milne, R. (2012). Predictors of eyewitness identification decisions from video lineups in England: A field study. *Law And Human Behavior, 36,* 257-265. doi:10.1037/h0093959

Howells, T. H. (1938). A study of ability to recognize faces. *The Journal Of Abnormal And Social Psychology, 33,* 124-127. doi:10.1037/h0062942

Innocence Project. (2010). Facts on post-conviction DNA exonerations. Retrieved from www.innocenceproject.org/know/

Jalbert, N. L., & Getting, J. (1992). Racial and gender issues in facial recognition. In F. Lösel, D. Bender, T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 309-316). Oxford England: Walter De Gruyter. doi:10.1515/9783110879773.309

Kahana, M. J., Rizzuto, D. S., & Schneider, A. R. (2005). Theoretical correlations and measured correlations: Relating recognition and recall in four distributed memory

models. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *31*, 933-953. doi:10.1037/0278-7393.31.5.933

Kassin, S. M., Tubb, V., Hosch, H. M., & Memon, A. (2001). On the 'general acceptance' of eyewitness testimony research: A new survey of the experts. *American Psychologist*, *56*, 405-416. doi:10.1037/0003-066X.56.5.405

Kebbell, M. R., & Milne, R. (1998). Police officers' perceptions of eyewitness performance in forensic investigations. *The Journal Of Social Psychology, 138*, 323-330. doi:10.1080/00224549809600384

Konečni, V. J., & Ebbesen, E. B. (1992). Methodological issues in research on legal decision-making, with special reference to experimental simulations. In F. Lösel, D. Bender, T. Bliesener (Eds.) , *Psychology and law: International perspectives* (pp. 413-423). Oxford England: Walter De Gruyter.

Kuehn, L. L. (1974). Looking down a gun barrel: Person perception and violent crime. *Perceptual And Motor Skills*, *39*, 1159-1164. doi:10.2466/pms.1974.39.3.1159

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Laney, C., & Loftus, E. F. (2009). Eyewitness memory. In R. N. Kocsis (Ed.), *Applied criminal psychology: A guide to forensic behavioral sciences* (pp. 121-145). Springfield, IL: Charles C Thomas Publisher.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: Women's faces make the difference. *Brain And Cognition, 50,* 121-128. doi:10.1016/S0278-2626(02)00016-7

Lewin, C., Wolgers, G., & Herlitz, A. (2001). Sex differences favoring women in verbal but not in visuospatial episodic memory. *Neuropsychology, 15*, 165-173. doi:10.1037/0894-4105.15.2.165

Lindholm, T., & Christianson, S. (1998). Gender effects in eyewitness accounts of a violent crime. *Psychology, Crime & Law, 4*, 323-339. doi:10.1080/10683169808401763

Lindsay, R. C., Lim, R., Marando, L., & Cully, D. (1986). Mock-juror evaluations of eyewitness testimony: A test of metamemory hypotheses. *Journal Of Applied Social Psychology, 16*, 447-459. doi:10.1111/j.1559-1816.1986.tb01151.x

Lindsay, R. L. (1994). Expectations of eyewitness performance: Jurors' verdicts do not follow from their beliefs. In D. Ross, J. Read, M. P. Toglia (Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 362-384). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511759192.018

Lindsay, R. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses: A replication and extension. *Law And Human Behavior, 13*, 333-339. doi:10.1007/BF01067033

Lloyd-Jones, T. J., & Brown, C. (2008). Verbal overshadowing of multiple face recognition: Effects on remembering and knowing over time. *European Journal Of Cognitive Psychology, 20*, 456-477. doi:10.1080/09541440701728425

Loftus, E. F., Miller, D. G., & Burns, H. J. (1987). Semantic integration of verbal information into a visual memory. In L. S. Wrightsman, C. E. Willis, S. M. Kassin (Eds.), *On the witness stand* (pp. 157-177). Thousand Oaks, CA: Sage Publications, Inc.

Malpass, R. S., & Kravitz, J. (1987). Recognition for faces of own and other race. In L. S. Wrightsman, C. E. Willis, S. M. Kassin (Eds.), *On the witness stand* (pp. 178-190). Thousand Oaks, CA: Sage Publications, Inc.

Manson v. Braithwaite, 432 U.S. 98 (1977).

McGivern, R. F., Mutter, K. L., Anderson, J., Wideman, G., Bodnar, M., & Huston, P. J. (1998). Gender differences in incidental learning and visual recognition memory: Support for a sex difference in unconscious environmental awareness. *Personality And Individual Differences*, *25*, 223-232. doi:10.1016/S0191-8869(98)00017-8

McKelvie, S. J. (1987). Sex differences, lateral reversal, and pose as factors in recognition memory for photographs of faces. *Journal of General Psychology, 114*, 13–37.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, And Law*, *7*, 3-35. doi:10.1037/1076-8971.7.1.3

Meissner, C. A., Brigham, J. C., & Kelley, C. M. (2001). The influence of retrieval processes in verbal overshadowing. *Memory & Cognition*, *29*, 176-186. doi:10.3758/BF03195751

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal Of Cognitive Psychology*, *20*, 414-455. doi:10.1080/09541440701728581

Memon, A., Gabbert, F., & Hope, L. (2004). The ageing eyewitness. In J. R. Adler (Ed.). *Forensic psychology: Concepts, debates and practice* (pp. 96-112). Devon United Kingdom: Willan Publishing.

Memon, A., Havard, C., Clifford, B., Gabbert, F., & Watt, M. (2011). A field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. *Psychology, Crime & Law*, *17*, 711-729. doi:10.1080/10683160903524333

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal Of Verbal Learning & Verbal Behavior*, *16*, 519-533. doi:10.1016/S0022-5371(77)80016-9

National Institute of Justice, Eyewitness Evidence: A Guide for Law Enforcement (1999).

National Institute of Justice, Eyewitness Evidence: A Trainer's Manual For Law Enforcement (October 2003).

Neil v. Biggers, 409 U.S. 188 (1972).

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science*, *16*, 65–69.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language*, *35*, 261-285. doi:10.1006/jmla.1996.0015

Pezdek, K. (2012). Fallible eyewitness memory and identification. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 105-124). Washington, DC: American Psychological Association. doi:10.1037/13085-005

Pigott, M. A., & Brigham, J. C. (1985). Relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology*, *70*, 547-555.

Pike, G., Brace N. & Kynan, S. (2002). The visual identification of suspects: procedures and practice. The visual identification of suspects: Procedures and Practice. *A Publication*

*of the Policing and Reducing Crime Unit, Home Office Research, Development and Statistics Directorate*.

Powers, P. A., Andriks, J. L., & Loftus, E. F. (1979). Eyewitness accounts of females and males. *Journal Of Applied Psychology, 64*, 339-347. doi:10.1037/0021-9010.64.3.339

Pozzulo, J. D., & Warren, K. L. (2003). Descriptions and identifications of strangers by youth and adult eyewitnesses. *Journal Of Applied Psychology, 88*, 315-323. doi:10.1037/0021-9010.88.2.315

Rehnman, J., & Herlitz, A. (2006). Higher face recognition ability in girls: Magnified by own-sex and own-ethnicity bias. *Memory, 14*, 289-296. doi:10.1080/09658210500233581

Rehnman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta Psychologica, 124*, 344-355. doi:10.1016/j.actpsy.2006.04.004

Ross, D. F., Read, J. D., & Toglia, M. P. (1994). *Adult eyewitness testimony: Current trends and developments*. New York, NY: Cambridge University Press.

Schooler, J. W. (2002). Verbalization produces a transfer inappropriate processing shift. *Applied Cognitive Psychology, 16*, 989-997. doi:10.1002/acp.930

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*, 36-71. doi:10.1016/0010-0285(90)90003-M

Schooler, J. W., Fiore, S. M., & Brandimonte, M. A. (1997). At a loss from words: Verbal overshadowing of perceptual memories. In D. L. Medin (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol. 37* (pp. 291-340). San Diego, CA: Academic Press.

Schooler, J. W., Ryan, R. S., & Reder, L. M. (1996). The costs and benefits of verbally rehearsing memory for faces. In D. Herrmann, M. K. Johnson, C. McEvoy, C. Hertzog, & P. Hertel (Eds.), Basic and applied memory: New findings (pp. 51-65). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Shapiro, P.N., Penrod, S. (1986). Meta-analysis of facial identification studies. *Psychological Bulletin, 10,* 139-156. doi:10.1037/0033-2909.100.2.139

Shaw, J. I., & Skolnick, P. (1994). Sex differences, weapon focus, and eyewitness reliability. *The Journal Of Social Psychology*, *134*, 413-420.

Shaw, J. I., & Skolnick, P. (1999). Weapon focus and gender differences in eyewitness accuracy: Arousal versus salience. *Journal Of Applied Social Psychology, 29,* 2328-2341. doi:10.1111/j.1559-1816.1999.tb00113.x

Sporer, S. L. (1996). Psychological aspects of person descriptions. In S. L. Sporer, R. S. Malpass, G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53-86). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Sporer, S. L., Malpass, R. S., & Koehnken, G. (1996). *Psychological issues in eyewitness identification*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

State v. Henderson, 209 N.J. 507; 39 A.3d 147 (2011).

State v. Lawson, 239 Or App 363, 244 P3d 860 (2012)

Steblay, N. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-298.

Tollerstrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. F. Ross, J. D. Read, M. P. Toglia

(Eds.), *Adult eyewitness testimony: Current trends and developments* (pp. 144-160). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511759192.008

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law And Human Behavior*, *21,* 661-685. doi:10.1023/A:1024812831576

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal Of Personality And Social Psychology, 36,* 1546-1557.

Wells, G. L. (1985). Verbal descriptions of faces from memory: Are they diagnostic of identification accuracy? *Journal of Applied Psychology, 70,* 619–626 doi:10.1037/0022-3514.36.12.1546

Wells, G. L., Greathouse, S. M., & Smalarz, L. (2012). Why do motions to suppress suggestive eyewitness identifications fail?. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 167-184). Washington, DC: American Psychological Association. doi:10.1037/13085-008

Wells, G. L., & Loftus, E. F. (1984). *Eyewitness testimony psychological perspectives*. London, England:Cambridge University Press.

Wells, G. L., & Murray, D. M. (1983). What can psychology say about the Neil v. Biggers criteria for judging eyewitness accuracy? *Journal of Applied Psychology, 68*, 347-362.

Wells, G. L., & Seelau, E. P. (1995). Eyewitness identification: Psychological research and legal policy on lineups. *Psychology, Public Policy, And Law, 1*, 765-791. doi:10.1037/1076-8971.1.4.765

Wells, G., Small, M., & Penrod, S. (1998) Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law & Human Behavior, 22*, 603-619.

Witryol, S. L., & Kaess, W. A. (1957). Sex differences in social memory tasks. *The Journal Of Abnormal And Social Psychology, 54*, 343-346. doi:10.1037/h0043411

Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition, 25,* 731-739. doi:10.3758/BF03211316

Wolfskiel, M. P., & Brigham, J. C. (1985). The relationship between accuracy of prior description and facial recognition. *Journal of Applied Psychology, 70*, 547-555. doi:10.1037/0021-9010.70.3.547

Woolnough, P. S., & MacLeod, M. D. (2001). Watching the birdie watching you: Eyewitness memory for actions using CCTV recording of actual crimes. *Applied Cognitive Psychology, 15*, 395-411. doi:10.1002/acp.717

Wright, D. B., & Davies, G. M. (1999). Eyewitness testimony. In F. T. Durso (Ed.). *Handbook of applied cognition* (pp. 789-818). New York, NY: John Wiley & Sons Ltd.

Wright, D. B. & McDaid, A. T. (1996). Comparing system and estimator variables using data from real lineups. *Applied Cognitive Psychology, 10*, 75-84.

Wright, D. B., & Sladden, B. (2003). An own gender bias and the importance of hair in face recognition. *Acta Psychologica, 114*, 101–114.

Yarmey, A. (1993). Adult age and gender differences in eyewitness recall in field settings. *Journal Of Applied Social Psychology*, *23*, 1921-1932. doi:10.1111/j.1559-1816.1993.tb01073.x

Yu, C. J., & Geiselman, R. E. (1993). Effects of constructing Identi-kit composites on photospread identification performance. *Criminal Justice And Behavior*, *20*, 280-292. doi:10.1177/0093854893020003005

Zelinski, E. M., Gilewski, M. J., & Schaie, K. (1993). Individual differences in cross-sectional and 3-year longitudinal memory performance across the adult life span. *Psychology And Aging*, *8*, 176-186. doi:10.1037/0882-7974.8.2.176

Table 1

*Description Accuracy Measure #1*

| | | | 4 = *very accurate* (matching actual raters' modal response) | | 3 = *slightly accurate* (a small difference from raters' modal response; or 1 SD away) | | 2 = *slightly inaccurate* (a moderate difference from raters' actual response; or 2 SDs away) | | 1 = *very inaccurate* (a large difference from raters' modal response; or 3 or more SDs away) | |
|---|---|---|---|---|---|---|---|---|---|---|
| Categories | Female | Male | Female | Male | Female | Male | Female | Male | Female | Male |
| 1. Race | White | White | White; Caucasian | White; Caucasian | - | - | - | - | Black | Black |
| 2. Sex | Female | Male | Female | Male | - | - | - | - | Male | Female |
| 3. Age (in years) | 24.63 | 27.67 | 24 - 25 | 27 - 28 | 22 - 23 or 26 - 27 | 25 - 26 or 28 - 29 | 20 - 21 or 28 - 29 | 23 - 24 or 30 - 31 | < 20 or > 29 | < 23 or > 31 |
| 4. Height (in inches) | 63.25 | 72.13 | 5'3" - 5'4" | 6'2" - 6'3" | 5'1" - 5'2" or 5'4" - 5'5" or average | 5'10" - 6'1" or 6'4" - 6'5" or average | 5'0" - 5'1" or 5'6" - 5'7" | 5'8" - 5'9" or 6'5" - 6'6" | < 5'0" or > 5'7" | < 5'8" or > 6'6" |
| 5. Weight (in lbs) | 131.86 | 182.29 | 131 - 132 lbs | 182 - 183 lbs | 120 - 130 lbs or 133 - 143 lbs average | 165 - 182 lbs or 183 - 199 lbs average | 109 - 119 lbs or 144 - 155 lbs | 149 - 165 lbs or 199 - 216 lbs | < 109 lbs or > 155 lbs | < 149 lbs or > 216 lbs |
| 6. Body build | Normal | Normal | Normal | Normal | Healthy | Healthy | Underweight, overweight | Underweight, overweight | Obese | Obese |
| 7. Hair color | Blonde | Brown | Blonde | Brown | Light, highlights | Black | Dark | Light | Any other color | Any other color |
| 8. Hair length | Long | Short | Long | Short | Medium | Medium | Short | Very short | Very short, bald | Very long, bald |
| 9. Jacket | Black | Black | Black | Black | Dark | Dark | Medium | Medium | Light | Light |
| 10. Shirt | Light | Light | Light | Light | Light color (any color) | Light color (any color) | Dark color | Dark color | Black | Black |
| 11. Pants | Shorts | Shorts | Shorts | Shorts | Denim shorts | Khaki shorts | Capris | Ankle pants | Long pants | Long pants |
| 12. Shoes | Sandals | Tennis shoes | Sandals | Tennis shoes | Barefoot | Shoes | Shoes | | Any footwear besides tennis shoes | Any footwear besides tennis shoes |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 13. Bag | Black bag | Black bag | Black bag, tote | Black bag, tote | Luggage bag | Luggage bag | Bagpack | Bagpack | Purse | Purse |
| 14. Lock pick | Lock pick | Lock pick | Lock pick (or other word for it) | Lock pick (or other word for it) | Weapon | Weapon | Stick | Stick | Specific stick (i.e. ballbat) | Specific stick (i.e. ballbat) |

*Note.* Responses for each description were divided into 14 categories and scored as follows: 4 = *very accurate* (matching raters' modal responses); 3 = *slightly accurate* (a small difference from raters' modal responses); 2 = *slightly inaccurate* (a moderate difference from raters' modal responses); and 1 = *very inaccurate* (a large difference from raters' modal responses). Description involving the actual measured characteristics of the criminals (age, height, and weight) were scored in terms of their degree of standard deviation from the correct response.

Table 2

*Description Accuracy Measures: Scoring Guide*

| | Directions |
|---|---|
| Description Accuracy Measure #1 | Responses for each description were divided into categories and scored as follows: 4 = *very accurate*, 3 = *slightly accurate*, 2 = *slightly inaccurate*, and 1 = *very inaccurate*. Descriptions involving the actual measured characteristics of the models (age, height, and weight) were scored in terms of the degree of standard deviation from the correct response. Factors not mentioned by a particular subject were not included in the calculation of his or her mean description accuracy score. |
| Description Accuracy Measure #2 | Mean description accuracy score (#1) multiplied by the number of traits used to describe the criminal. |
| Description Accuracy Measure #3 | Similar to #1, except that subject's nonresponses were treated as zeros, making a subject's mean score lower because they would have omitted a description. |
| Description Accuracy Measure #4 | This treated responses as either right or wrong. For a subject's response to be correct, it had to exactly match the rater's mean or modal response for that particular characteristic. Correct responses were scored as +1, while all other responses were scored as 0. The highest score could have been 14 (because we have 14 categories). |
| Descriptor Quantity | Total number of descriptors/adjectives. |

*Note.* This scoring guide was taken directly from Pigott and Brigham (1985). However, we included and defined the descriptor quantity measure. This guide directed the research assistants' scoring of descriptions.

Table 3

*Correlation Matrix for Description Accuracy and Descriptor Quantity Measures*

| | Description Accuracy Measure #1 | Description Accuracy Measure #2 | Description Accuracy Measure #3 | Description Accuracy Measure #4 | Descriptor Quantity |
|---|---|---|---|---|---|
| Description Accuracy Measure #1 | | | | | |
| *r* | 1 | .33* | .43* | .50* | -0.03 |
| Description Accuracy Measure #2 | | | | | |
| *r* | .33* | 1 | .75* | .60* | .93* |
| Description Accuracy Measure #3 | | | | | |
| *r* | .43* | .75* | 1 | .80* | .62* |
| Description Accuracy Measure #4 | | | | | |
| *r* | .50* | .60* | .80* | 1 | .43* |
| Descriptor Quantity | | | | | |
| *r* | -0.03 | .93* | .62* | .43* | 1 |

*Note*. Each of the four measures of description accuracy was correlated. The relationships between all four measures of description accuracy were found to be significantly correlated with each other. Three of the description accuracy measures were found to be significantly correlated with the measure of descriptor quantity. Significant correlations are indicated with a *.

* $p < .001$

Table 4

*Description Accuracy and Descriptor Quantity Measures by Participant Gender*

| | *M* | *SD* | Minimum | Maximum | Range |
|---|---|---|---|---|---|
| Description Accuracy Measure #1 | | | | | |
| Males | 3.23 | 0.36 | 2.25 | 4.00 | 1.75 |
| Females | 3.27 | 0.39 | 1.40 | 4.00 | 2.60 |
| Total | 3.26 | 0.38 | 1.40 | 4.00 | 2.60 |
| Description Accuracy Measure #2 | | | | | |
| Males | 22.89 | 6.96 | 9.75 | 41.14 | 31.39 |
| Females | 25.75* | 8.22 | 6.00 | 51.43 | 45.43 |
| Total | 24.76 | 7.91 | 6.00 | 51.43 | 45.43 |
| Description Accuracy Measure #3 | | | | | |
| Males | 1.33 | 0.41 | 0.43 | 2.36 | 1.93 |
| Females | 1.45 | 0.47 | 0.21 | 2.43 | 2.22 |
| Total | 1.41 | 0.45 | 0.21 | 2.43 | 2.22 |
| Description Accuracy Measure #4 | | | | | |
| Males | 3.41 | 1.88 | 0.00 | 16.00 | 16.00 |
| Females | 3.58 | 1.62 | 0.00 | 8.00 | 8.00 |
| Total | 3.52 | 1.72 | 0.00 | 16.00 | 16.00 |
| Descriptor Quantity | | | | | |
| Males | 7.09 | 2.01 | 3.00 | 12.00 | 9.00 |
| Females | 7.89* | 2.37 | 2.00 | 15.00 | 13.00 |
| Total | 7.61 | 2.28 | 2.00 | 15.00 | 13.00 |

*Note.* The means, standard deviations, and ranges of each of the four description accuracy measures and descriptor quantity measure are displayed in the table. The significant main effects for participant gender (female superiority main effect) are indicated with a *, which means that female participants had significantly higher ($p \leq .05$) description accuracy scores and descriptor quantity scores compared to male participants.

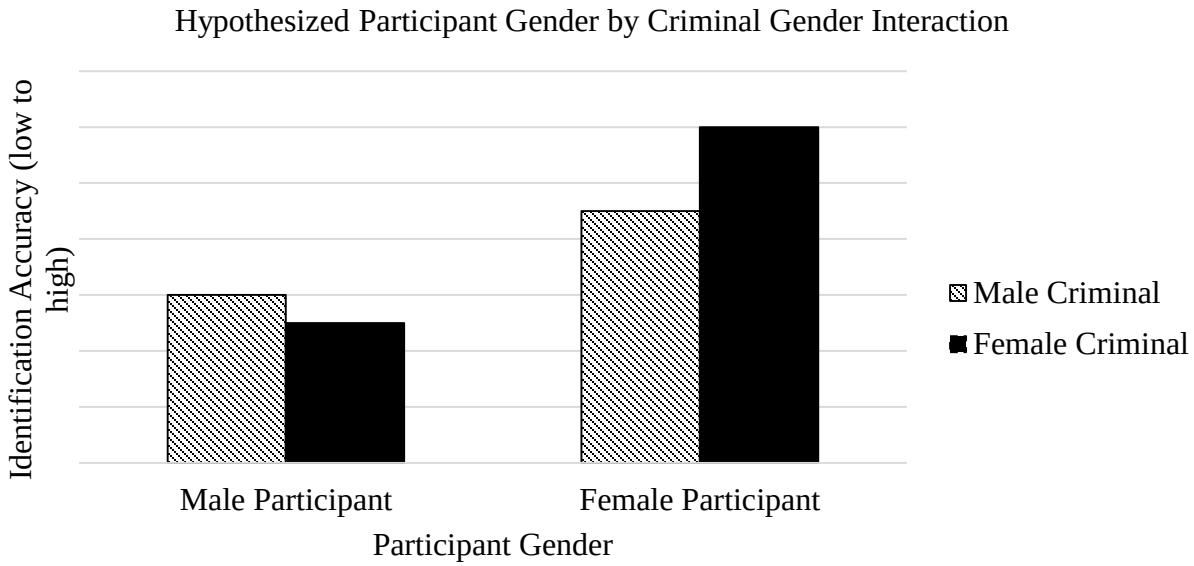



*Figure 1.* Hypothesized participant gender by criminal gender interaction. Female eyewitnesses will recognize and correctly identify female culprits in a lineup better than male eyewitnesses will recognize and correctly identify female culprits.

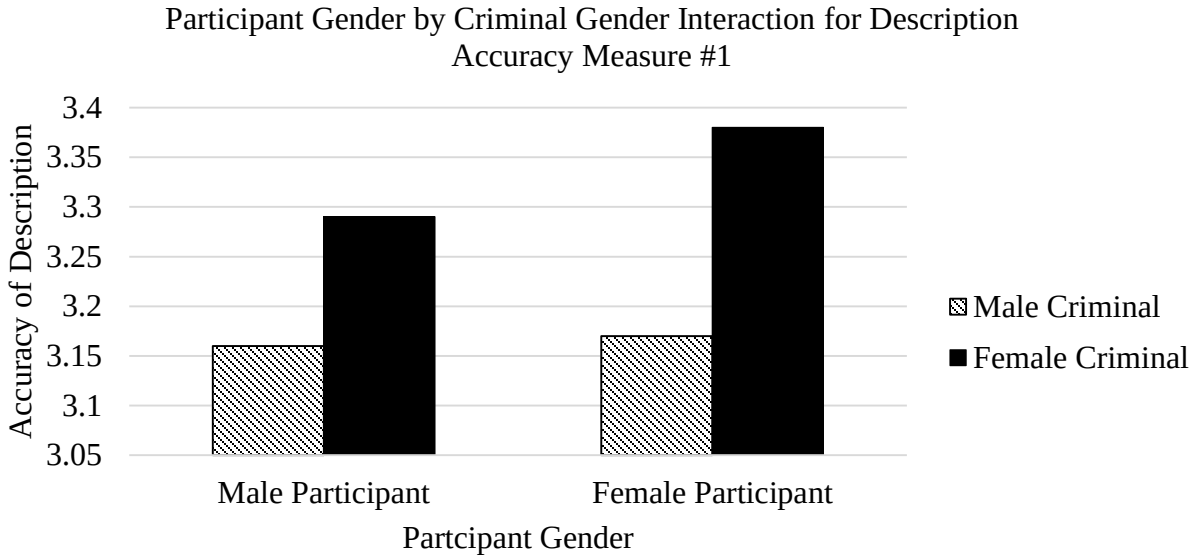



*Figure 2*. Participant gender by criminal gender interaction for description accuracy measure #1.

# Appendix A

## Notice of IRB Approval

**From:** Dr. Stan Aeschleman, Institutional Review Board Chairperson

**Date:** 2/17/2014

**RE:** Notice of IRB Approval by Expedited Review (under 45 CFR 46.110)

**Study #:** 14-0055

**Study Title:** Crime of Breaking and Entering

**Submission Type:** Modification

**Expedited Category:** (7) Research on Group Characteristics or Behavior, or Surveys, Interviews, etc.

**Approval Date:** 2/17/2014

**Expiration Date of Approval:** 9/25/2014

The Institutional Review Board (IRB) approved the modification for this study. The IRB found that the research procedures meet the expedited category cited above. IRB approval is limited to the activities described in the IRB approved materials, and extends to the performance of the described activities in the sites identified in the IRB application. In accordance with this approval, IRB findings and approval conditions for the conduct of this research are listed below.

**Submission Description:**

Data collection has no occurred yet, but we have decided to go with a U.S. sample versus a student sample.

Changes made to consent form in regards to Amazon Mech-Turk and amount of compensation.

Changes made to application in regards to delayed consent, number of participants sought,

and participants sought (>18).

**Regulatory and other findings:**

The IRB waived the requirement to obtain a signed consent form for some or all subjects because the research presents no more than minimal risk of harm to subjects and involves no procedures for which written consent is normally required outside of the research context.

**Approval Conditions:**

Appalachian State University Policies: All individuals engaged in research with human participants are responsible for compliance with the University policies and procedures, and IRB determinations.

Principal Investigator Responsibilities: The PI should review the IRB's list of PI responsibilities. The Principal Investigator (PI), or Faculty Advisor if the PI is a student, is ultimately responsible for ensuring the protection of research participants; conducting sound ethical research that complies with federal regulations, University policy and procedures; and maintaining study records.

Modifications and Addendums: IRB approval must be sought and obtained for any proposed modification or addendum (e.g., a change in procedure, personnel, study location, study instruments) to the IRB approved protocol, and informed consent form before changes may be implemented, unless changes are necessary to eliminate apparent immediate hazards to participants. Changes to eliminate apparent immediate hazards must be reported promptly to the IRB.

Approval Expiration and Continuing Review: The PI is responsible for requesting continuing review in a timely manner and receiving continuing approval for the duration of the research with human participants. Lapses in approval should be avoided to protect the welfare of

enrolled participants. If approval expires, all research activities with human participants must cease.

Prompt Reporting of Events: Unanticipated Problems involving risks to participants or others; serious or continuing noncompliance with IRB requirements and determinations; and suspension or termination of IRB approval by external entity, must be promptly reported to the IRB.

Closing a study: When research procedures with human subjects are completed, please complete the Request for Closure of IRB review form and send it to irb@appstate.edu.

Appendix B

Informed Consent Forms

**Consent to Participate in Research:** *Information to Consider About this Research*

**Crime of Breaking and Entering Part 1**

Principal Investigator: Melissa Baker bakerma@appstate.edu, Bethany Poff poffba@appstate.edu, Dr. Paul Fox (faculty advisor) foxpa@appstate.edu.

You are being invited to take part in a research study about verbal ability and participant gender. You will be asked to view a brief video, asked a series of questions, and will complete an anagram task. This study will not take longer than 30 minutes. This study offers $0.20. This study has two parts. If you choose to participate in part 2 you must sign up and do so 1-4 days later upon completion of part 1. For participation in part 2 you will receive an additional $0.50. Credit for part 1 of the study does not depend on participation in part 2 of the study.

You cannot volunteer for this study if you are under 18 years of age.

**What are the possible benefits and risks of the research?**
There may be no personal benefit from your participation but the information from this research may help others in the future by learning about verbal ability and gender of participant.

To the best of our knowledge, the risk of harm for participating in this research study is no more than you would experience in everyday life. If you find some of the questions we ask to be upsetting or stressful, please contact the Appalachian State University Counseling and Psychological Services, Monday-Friday 8:00am-5:00pm, (828)-262-3180.

You will be asked to sign the consent form electronically in order to give you course credit. The PI will separate your name from the data. No one other than the members of the research team will be able to associate the name to the data.

Your information will be combined with information from other people taking part in the study. When we write up the study to share it with others, we will write about the combined information. You will not be identified in any published or presented materials. We will protect your confidentiality by dissociating your name from the data after course credit and a code is assigned. The data and identifying information will be securely stored electronically for three years.

**Who can I contact if I have questions?**
The people conducting this study will be available to answer any questions concerning this research, now or in the future. You may contact the Principal Investigator at bakerma@appstate.edu, poffba@appstate.edu, or foxpa@appstate.edu.

**Question regarding the protection of human subjects in research projects can be directed to the IRB Administrator:**
Research and Sponsored Programs
Appalachian State University
Boone, NC 28608
(828)-262-2130
irb@appstate.edu

The research has been approved on 09-26-2013 by the Institutional Review Board (IRB) at Appalachian State University. This approval will expire on 09-25-2014 unless the IRB renews the approval of this research.

**Do I have to participate? What else should I know?**
Your participation in this research is completely voluntary. There will be no consequences if you choose not to volunteer and may decide to stop participating at any time. If you decide not to participate in Part 2 of the study, no penalty will be enforced. You will receive $0.20 for Part 1.

**I have decided I want to take part in this research. What should I do now?**

Please read the following and if you agree, you should indicate your agreement by entering your name in the space provided and clicking next to begin the survey:

1. I have read all of the above information.
2. I understand that I can stop taking part in this study at any time.
3. I understand that I am not giving up any of my rights.
4. I may obtain a copy of this consent form by contacting the principal investigator as listed above.

_____

Participant's Name Date

**Consent to Participate in Research:** *Information to Consider About this Research*

**Crime of Breaking and Entering Part 2**

Principal Investigator: Melissa Baker bakerma@appstate.edu, Bethany Poff poffba@appstate.edu, Dr. Paul Fox (faculty advisor) foxpa@appstate.edu.

You are being invited to take part in a research study about a person breaking and entering into a home. By doing this study we hope to learn about eyewitness testimony. You will be asked to view a brief video and will answer questions about the video that you saw in part 1 of the study, and answer questions. You will be shown a series of photos and asked to identify the perpetrator in the video if he/she is present. This study will not take longer than 30 minutes. You will receive additional $0.50 for participating in this part. You may not participate in part 2 if you have not already participated in part 1 of this study.

You cannot volunteer for this study if you are under 18 years of age.

**What are the possible benefits and risks of the research?**
There may be no personal benefit from your participation but the information from this research may help others in the future by learning about breaking and entering cases.

To the best of our knowledge, the risk of harm for participating in this research study is no more than you would experience in everyday life. If you find some of the questions we ask to be upsetting or stressful, please contact the Appalachian State University Counseling and Psychological Services, Monday-Friday 8:00am-5:00pm, (828)-262-3180.

You will be asked to sign the consent form electronically in order to give you course credit. The PI will separate your name from the data. No one other than the members of the research team will be able to associate the name to the data.

Your information will be combined with information from other people taking part in the study. When we write up the study to share it with others, we will write about the combined information. You will not be identified in any published or presented materials. We will protect your confidentiality by dissociating your name from the data after course credit and a code is assigned. The data and identifying information will be securely stored electronically for three years.

**Who can I contact if I have questions?**
The people conducting this study will be available to answer any questions concerning this research, now or in the future. You may contact the Principal Investigator at bakerma@appstate.edu, poffba@appstate.edu, or foxpa@appstate.edu.

**Question regarding the protection of human subjects in research projects can be directed to the IRB Administrator:**
Research and Sponsored Programs
Appalachian State University

Boone, NC 28608
(828)-262-2130
irb@appstate.edu

The research has been approved on 09-26-2013 by the Institutional Review Board (IRB) at Appalachian State University. This approval will expire on 09-25-2014 unless the IRB renews the approval of this research.

**Do I have to participate? What else should I know?**
Your participation in this research is completely voluntary. There will be no consequences if you choose not to volunteer and may decide to stop participating at any time. If you decide not to participate in Part 2 of the study, no penalty will be enforced. You will receive $0.20 for Part 1.

**I have decided I want to take part in this research. What should I do now?**

Please read the following and if you agree, you should indicate your agreement by entering your name in the space provided and clicking next to begin the survey:

5. I have read all of the above information.
6. I understand that I can stop taking part in this study at any time.
7. I understand that I am not giving up any of my rights.
8. I may obtain a copy of this consent form by contacting the principal investigator as listed above.

_____

Participant's Name Date

# Appendix C

## Crime Videos

### Female Crime Video

Now you will be shown a brief video. The video does not require sound. The video will begin automatically after you press <next>. You may not pause, rewind, or replay the video. Pay attention. You will be asked about the video later.


Male Crime Video

Now you will be shown a brief video. The video does not require sound. The video will begin automatically after you press <next>. You may not pause, rewind, or replay the video. Pay attention. You will be asked about the video later.


# Appendix D

## Anagram Task

Table 1. *Anagrams and Anagrams Solutions.*

| Practice Anagrams | | |
|---|---|---|
| | Solutions | Anagrams |
| 1. | BLIMP | PLIMB |
| 2. | CLOWN | OWNLC |
| 3. | MONTH | HTMON |

| Anagrams | | |
|---|---|---|
| 1. | APRON | ONAPR |
| 2. | BUNCH | BNHUC |
| 3. | CABIN | NBIAC |
| 4. | CHAIR | CIAHR |
| 5. | DOUBT | UTDBO |
| 6. | FLASH | HSAFL |
| 7. | FRUIT | ITRUF |
| 8. | GLOVE | GEVOL |
| 9. | HOUND | HNDUO |
| 10. | JOINT | IJNOT |
| 11. | JUDGE | JEGUD |
| 12. | KNIFE | FNKIE |
| 13. | LOGIC | IOGLC |
| 14. | MAJOR | OAJRM |
| 15. | MERCY | EMCYR |
| 16. | PLANK | LAKPN |
| 17. | PORCH | OCPHR |
| 18. | SCOUT | OUSTC |
| 19. | SNACK | KASNC |
| 20. | STYLE | TELSY |
| 21. | TRUCK | KRTCU |
| 22. | UNITY | IUNYT |
| 23. | VAULT | AVTLU |
| 24. | WALTZ | ZLTWA |
| 25. | WOMAN | OWAMN |

*Note*. Above are the 25 anagrams used from Gilhooly and Johnson (1978) analysis of anagram difficulty. Participants were told, “You will be shown several anagrams. Each anagram contains 5 letters. You will have 15 seconds to solve each anagram."

## Appendix E

## Demographic Information Questionnaire

Please note, your information will not be given to outside entities. It is for internal use only.

1. What is your age?

2. What is your sex?

   Male
   Female

3. What is your current marital status?
   Single
   Married
   Separated
   Divorced
   Widowed

4. How do you describe yourself? (Please check the one option that best describes you.)

   Black or African American
   Hispanic or Latina
   Non-Hispanic White
   Other [fill in]

5. What state do you live in?

6. What is the highest level of education you completed?
   Elementary school only
   Some high school, but did not finish
   Completed high school
   Some college, but did not finish
   Two-year college degree / A.A. / A.S.
   Four-year college degree / B.A. / B.S.
   Some graduate work
   Completed Master's or professional degree
   Advanced graduate work or Ph.D.

7. What is your employment status?
   Employed full-time
   Employed part-time
   Unemployed
   Student

Homemaker
Retired

8. What is your annual income?

   Under $25,000
   $25,000 - $39,999
   $40,000 - $49,999
   $50,000 - $74,999
   $75,000 - $99,999
   $100,000 - $124,999
   $125,000 - $149,999
   Over $150,000

9. How long have you been participating in Amazon Mechanical Turk surveys?
   Less than 1 year
   1 – 2 years
   3 – 4 years
   5 – 6 years
   More than 6 years

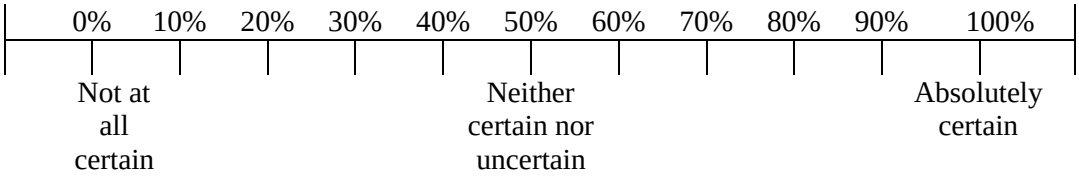
10. How many hours per week do you spend on Amazon Mechanical Turk participating in surveys?
    0 – 2 hours per week
    2 – 4 hours per week
    4 – 6 hours per week
    6 – 8 hours per week
    More than 8 hours per week

# Appendix F

## Confidence, Clarity of Memory, and Distraction Rating Likert Scales

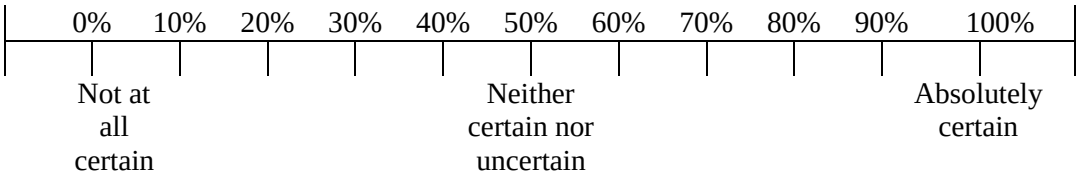
### Pre-confidence Rating

Indicate the percent of certainty you have in your ability to identify the criminal you saw in the previous video from a lineup if the criminal was present.



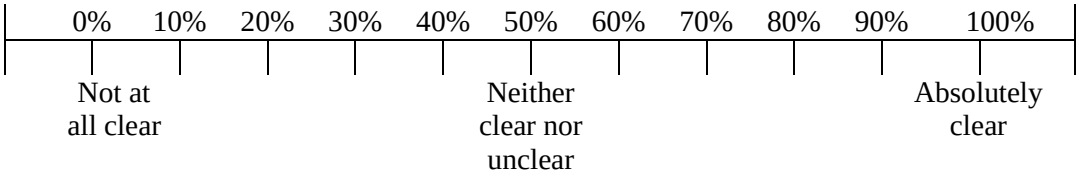
Post-confidence Rating

Indicate the percent of certainty that you made the correct decision when you selected the perpetrator shown in the lineup.



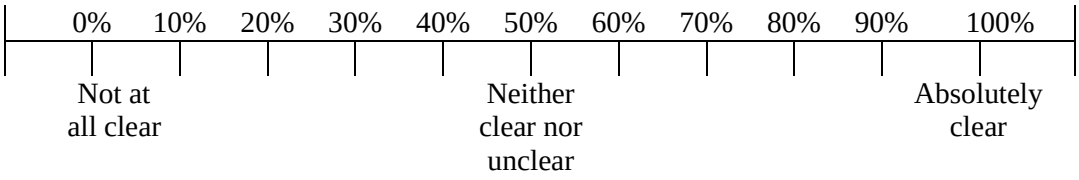
Pre-clarity of Memory Rating

Indicate your clarity of memory that you have the criminal that you san in the previous video.



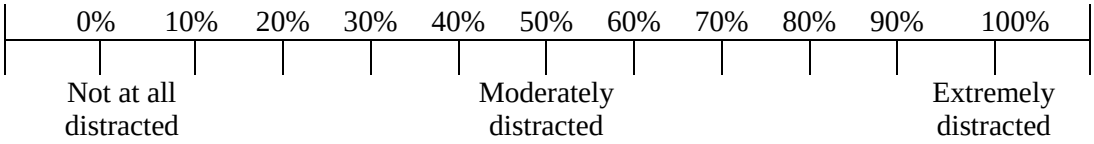
Post-clarity of Memory Rating

Indicate your clarity of memory that you have the criminal that you saw in the previous video.



Distraction Rating

To what extent were you distracted by your surroundings while participating in this study?



# Appendix G

## Description Task

Please provide a description of the criminal that you saw in the video at the beginning of the study. You should include a description of the criminal's physical appearance (examples include height, weight, hair style, etc.).

# Appendix H

## Lineups

### Female Lineup

If you believe that the person who committed the burglary is in the lineup, indicate your choice using the numbers above or below that person's photo (1-6). If you believe that the person is not in the lineup, you may choose 7.

*Note*. The identification task was completed by participants either indicating a photo in the lineup or selecting "none of them" if they believe the target was absent. Participants were randomly assigned to see the photo of the target in either positions 1-6.

*The target is in position 4 in this lineup.

Male Lineup

If you believe that the person who committed the burglary is in the lineup, indicate your choice using the numbers above or below that person's photo (1-6). If you believe that the person is not in the lineup, you may choose 7.

*Note*. The identification task was completed by participants either indicating a photo in the lineup or selecting "none of them" if they believe the target was absent. Participants were randomly assigned to see the photo of the target in either positions 1-6.

*The target is in position 2 in this lineup.

## Vita

Melissa Ann Baker was born July 10, 1991, in Charlotte, North Carolina, to Nancy and Allen Baker. In June 2009, she graduated from Bessemer City High School in Bessemer City, North Carolina. The following fall, August 2009, she attended Appalachian State University. In August 2012, she acquired her Bachelors of Science in Psychology with a minor in Criminal Justice, as well as a passion for doing research. After graduation, Melissa decided to further her research experience and completed an internship in Sydney, Australia, at the University of New South Wales and Armchair Psychology Inc. In August 2013, Melissa began study toward a Masters of Arts in General Experimental Psychology. She earned the degree in August 2015 as well as valuable academic and research skills. Melissa has presented her research at two national conferences and will continue to research various topics in the field of forensic psychology. In the spring of 2015, Melissa accepted an adjunct faculty member position at Appalachian State University, where she will teach undergraduate psychology courses, continue to do research, and pursue attainment of a Ph.D.